# A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States

**Paul Sutton, Dar Roberts, Chris Elvidge, and Henk Meij**

## Abstract

*The striking apparent correlation between nighttime satellite imagery and human population density was explored for the continental United States. The nighttime stable-lights imagery was derived from the visible near-IR band of 231 orbits of the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS). The population density data were generated from a gridded vector dataset of the 1992 United States census block group polygons. Both datasets are at a one-square-kilometre resolution. The two images were co-registered and correlation between them was measured at a range of spatial scales, including aggregation to state and county levels. DMSP imagery showed strong correlations at aggregate scales, and analysis of the saturated areas of the images showed strong correlations between the areas of saturated clusters and the populations those areas cover. The non-zero pixels of the DMSP imagery correspond to only 10 percent of the land cover yet account for over 80 percent of continental United States population. Spatial analysis of the clusters of the saturated pixels predicts population with an $R^2$ of 0.63. Consequently, the DMSP imagery may prove to be useful to inform a "smart interpolation" program to improve maps and datasets of human population distributions in areas of the world where good census data may not be available or do not exist.*

## Introduction

The growth in human population has profound social, economic, and environmental consequences. Accurate data on the spatial distribution of human population is critical in addressing the causes and impacts of global environmental change. Information concerning the arrangement of human population might improve proactive responses to the environmental degradation that often accompanies high population densities. Knowledge of human settlement trends at a global scale is essential in order to manage the rapid and inevitable urbanization of the planet (Tolba, 1992). According to Ehrlich (1988), the primary cause of the loss of biodiversity is the habitat destruction resulting from the expansion of human populations and human activities. The global effects of land-cover conversion on ecosystems and human wealth and well-being may be much larger than those arising from climate change (Skole, 1994). High quality data on the size and distribution of the human population over the whole planet is critical in order to monitor, understand, respond to, and perhaps even prevent environmental degradation, loss of biodiversity, and resource depletion in many parts of the world.

Nonetheless, consistent population data useful for these purposes does not exist (see Clark and Rhind (1992) for an excellent survey of global demographic data). One of the most comprehensive global demographic datasets was recently compiled at the University of California at Santa Barbara by the National Center for Geographic Information and Analysis (NCGIA). This global demography project was a joint effort of the NCGIA and the Consortium for International Earth Science Information Network (CIESIN) with additional funding from the Environmental Systems Research Institute (ESRI) (Tobler *et al.*, 1995). While this project serves an important function in addressing the need for data on the human population, it is limited in its spatial resolution and is extremely difficult to update. The global demography dataset was produced by gathering available census data from all the nations of the world. The level of aggregation was the second sub-national administrative unit (e.g., U.S. counties). However, data at this level of aggregation were not always available. For example, the only data available for Saudi Arabia was one national figure (Gottsegen, 1995, personal communication). In addition, the data that were provided by the various nations of the world may be very suspect. For example, a study of the migration data for the sub-national administrative units in China suggests that the one-child policy has caused a significant underestimate of the population of China (Deng, 1994). Instead of the conventional estimate of 1.2 billion for the population of China, Deng asserts it could be as high as 1.5 billion. The difference between these two estimates is greater than the total population of the United States.

This paper describes the comparison of two datasets that cover the continental United States. The first image is a 1-$km^2$ resolution grid of the population density derived from the 1990 United States decennial census. The grid was derived from the block group layer of the Bureau of the Census Topologically Integrated Geo-referenced and Encoded Referencing System (TIGER), and proportionally allocated to 1-$km^2$

P. Sutton and D. Roberts are with the Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106 (sutton@geog.ucsb.edu).

C. Elvidge is with the Desert Research Institute and Nevada Agricultural Experiment Station, University of Nevada System, Reno, NV 89506.

Meij is with the Socioeconomic Data and Applications Center (SEDAC), Consortium for International Earth Science Information Network (CIESIN), 2250 Pierce Road, University Center, MI 48710.

Photogrammetric Engineering & Remote Sensing, Vol. 63, No. 11, November 1997, pp. 1303–1313.

0099-1112/97/6311-1303$3.00/0

cells. This dataset was developed by the Socioeconomic Data and Applications Center (SEDAC) at CIESIN (Meij, 1995). The second dataset was produced by Elvidge *et al.* (1995) at the NOAA National Geophysical Data Center. This dataset is a composite image of 231 orbits of the nighttime passes of the Defense Meteorological Satellite Program (DMSP) over the continental United States. These two images were co-registered and compared at various scales and by using several methods of spatial aggregation.

## Background

The use of remote sensing techniques as a means of estimating human population parameters is not new; however, previous methods have been site-specific or computationally intensive, which makes them inadequate for generalization to a global scale (Forster, 1985). Ogrosky (1975) achieved a high degree of correlation (0.96) between population and the logarithm of image area classified as urban in the Puget Sound area; however, the nature of these relationships have been shown to vary from region to region. In addition, the classification scheme for defining "urban" is likely to vary from region to region. The work of Clayton and Estes (1980) was a check on census enumeration accuracy in the Goleta Valley and involved the counting of buildings in high altitude color infrared photographs. This method is clearly too labor intensive to apply at larger scales.

The regionally unique nature of human settlement patterns makes it extremely difficult to generalize any findings beyond the regions of inquiry. The signal detected in daytime imagery is produced primarily by reflected sunlight, most of which is not influenced by human settlement or activity; in contrast, most of the VNIR radiation detected in the stable lights DMSP OLS image is produced by human activities. Consequently, nighttime visible and near-infrared emissions may prove to be a robust proxy for human population. In addition, the use of ancillary socio-economic data by employing a geographic information system (GIS) may prove to be a means of correcting for any site-specific regional variation. This method has the potential to improve our estimates of population parameters, particularly for regions where current data are lacking.

Nighttime imagery provided by the Defense Meteorological Satellite Program (DMSP-OLS) has been available since the early 1970s. The DMSP sensors are more than four orders of magnitude more sensitive to visible near-infrared radiances than traditional satellite sensors optimized for daytime observation (Elvidge *et al.*, 1995). Observations of the striking qualitative correlation between DMSP imagery and maps of population distribution have undoubtedly motivated many studies (Croft, 1977; Croft, 1978; Foster, 1991) (see Figures 1a and 1b). Simple quantitative correlations between light levels and population density have not been identified. However, other indirect means of identifying strong quantitative relationships between satellite imagery and population have been explored.

Tobler (1969) used satellite imagery to confirm settlement size coefficients for the equation $r = a \times P^b$, where $r$ is the radius of the of the populated circle, $a$ is an empirically derived constant of proportionality, $P$ is the population, and $b$ is an empirically derived exponent. Estimates of these parameters are fairly consistent at regional scales but the estimate of the $a$ parameter vary markedly between regions (Boyce, 1963; Maher, 1969; Nordbeck, 1965; Stewart, 1958). DMSP OLS imagery has also been correlated with energy consumption, $r = a \times X^b$, where $X$ is energy rather than population, producing a correlation coefficient of 0.89 (Welch, 1980). Another study by Welch and Zupko (1980) used densitometry methods on older DMSP imagery which were generated in an analog manner on mylar films. In this study, correlation coefficients of 0.95 and 0.96 were found between the DMSP imagery and population of cities utilizing the same formula, $r = a \times P^b$. Energy consumption data were also obtained for the cities investigated and correlated with the DMSP imagery in a similar manner.

One drawback of these approaches is the fact that the $a$ parameter varies substantially across the globe, despite the fact that it is consistent at regional scales. Clearly, there are cultural, economic, and/or environmental determinants of this parameter that cannot be obtained from the satellite imagery alone. The existing findings show a spatial variation in this parameter; however, if there are other determinants of this parameter, it is likely that it will also vary with time. This suggests that identifying known spatial and possible temporal variation in these parameters may provide a method for developing a systematic and operationalizable means of using DMSP imagery to model urban populations at regional scales. One area of future research may determine if there are systematic means for extrapolating these methods across regions by utilizing national aggregate data such as percent of population in urban areas, GDP per capita, energy consumption per capita, etc. The early investigations were hampered by the nature of the data itself. Digital DMSP data were not archived until 1992. The resolution of the DMSP imagery was coarse, and the computing power needed to filter out clouds and fires was not available even if the data were available in digital format. Consequently, the investigations focused on specific urban areas, and most of the land areas of the regions investigated were ignored. Improvements in computing power, sophisticated GIS and image processing software, and the availability of the data in digital format suggest a re-examination of these techniques and their potential for monitoring and/or modeling the human population using nighttime satellite imagery is worthy of investigation.

## Methods

The development of the digital DMSP archive has dramatically improved access to and utility of the DMSP data. DMSP data are now available in digital format, and algorithms developed by Elvidge *et al.* (1997) have produced a 1-km-square resolution dataset of the city lights of the continental United States. Elvidge *et al.* developed algorithms to identify spatially stable VNIR emission sources utilizing hundreds of orbits and the infrared band of the DMSP system to screen out cloud impacted data.

At the time this research was conducted, there were two versions of these data available. In one version of the data, the value in the pixel was a percentage of times light was seen relative to the number of cloud free orbits for which that pixel was sampled. The other version simply used the maximum light level for a pixel for those cloud free orbits. The data used for this analysis were the maximum value per pixel. The rationale for choosing the maximum rather than the percent version of these datasets is because the physical interpretation is more direct and it proved more amenable to the subsequent spatial cluster analysis that was performed. The percent data is nonetheless quite interesting because there is a greater degree of variability in the pixel values within the urban clusters, which may prove to be a better dataset for identifying a direct correlation between pixel value and population density. However, since this analysis was performed, another dataset has been produced from the low-gain DMSP OLS system of the nighttime city lights. This dataset has much more variability within the urban clusters and has a more direct physical interpretation for using light intensity to predict population density on a pixel-by-pixel basis. This avenue is presently being explored. For a detailed

(a)

(b)

Figure 1. (a) Population density of the United States derived from the 1990 U.S. Census. (b) Composite DMSP image of the Continental United States.

discussion of these methods, see the aforementioned reference. The Elvidge *et al.* (1997) data were coregistered with the dataset developed by SEDAC at CIESIN.

Two simple tests were run to ensure that the values of the population density pixels were reasonable and that the co-registration and geolocation of the images were reasonable. The tests involved the use of vector GIS coverages of the continental United States being overlaid over each image. The population density pixels were then integrated (the DN of each pixel was summed for all the pixels in each state) and compared with the published census population of each state. This resulted in a set of 49 pairs of population points. A regression was run on these points and resulted in an $R^2$ of 0.999. A similar test was run on the DMSP OLS stable lights image in which the count of the number of pixels per state was compared to the known area of each state. The $R^2$ for

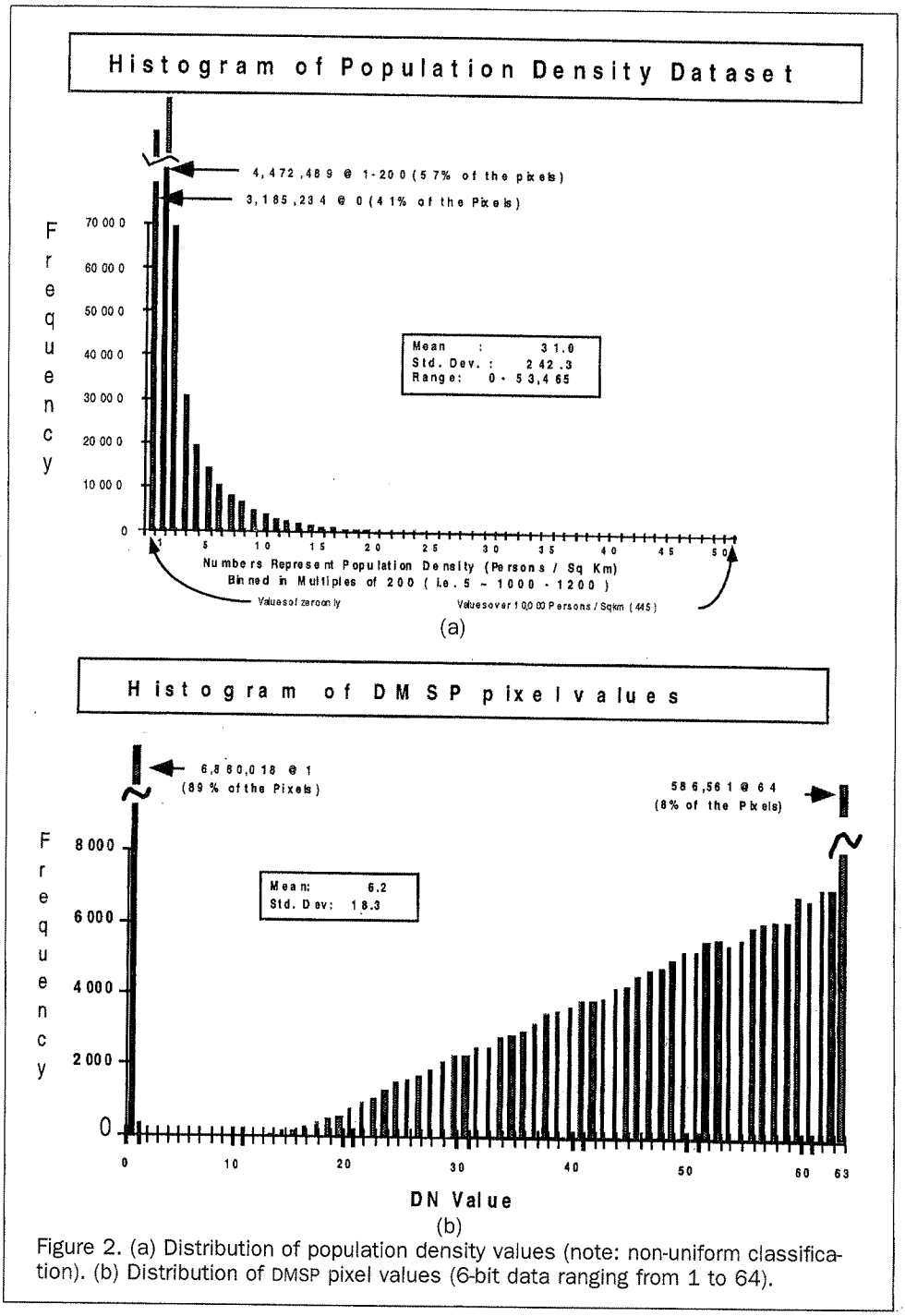



Figure 2. (a) Distribution of population density values (note: non-uniform classification). (b) Distribution of DMSP pixel values (6-bit data ranging from 1 to 64).

this regression was 0.999 also. The slopes of these regression lines were not significantly different from one. Similar regressions were run on data generated by aggregating to a vector coverage of the United States counties with similarly high correlations.

A focus of this investigation was to quantify the correlation between these two images at a range of scales and to explore various aggregation techniques and data transformations to maximize the correlation. The DMSP OLS dataset is available in an Interrupted Goode's Homolosine projection. The re-projection of this dataset into the Lambert Azimuthal Equal Area (LAZEA) projection is shown in Figure 1a. The population dataset is also available in the LAZEA projection (Figure 1a). The DMSP OLS stable lights dataset was re-projected to the projection of the population density dataset for the purpose of quantifying any correlations between the two images.

The DMSP OLS data have 6-bit quantization, providing a dynamic range between 0 and 63. Once the stable lights image was geo-referenced, the land values were incremented by one unit to distinguish land pixels from ocean pixels. Of the pixels that are part of the land of the continental United States, 89 percent have a value of one, 8 percent have a value of 64 (saturated), and 3 percent have intermediate values from 2 to 63 (Figure 2b). It should be noted that the frequency of all the intermediate values monotonically increases from low values to high. Clearly, many of the pixels in urban areas are saturated at 64, and this presents many

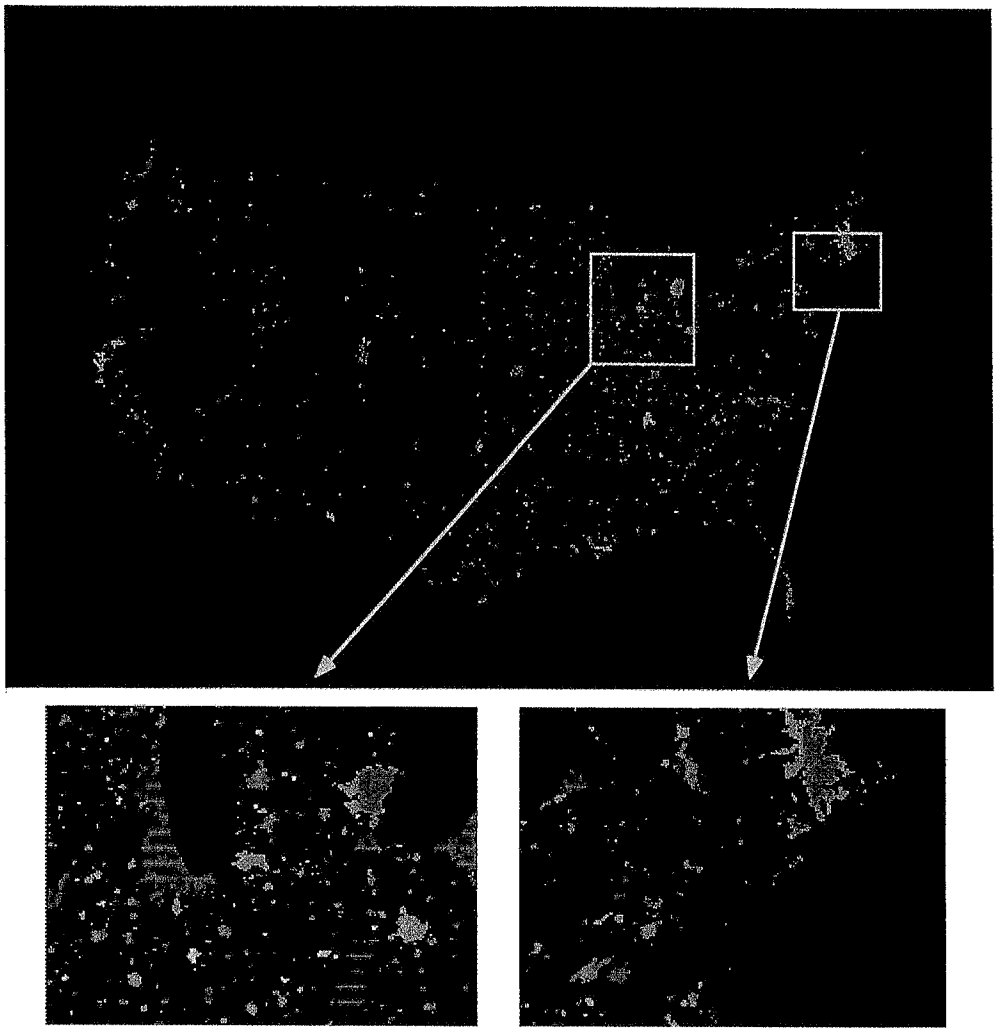


Plate 1. Saturated DMSP pixels classed as Urban "clusters" based on adjacency with insets of the Chicago-Detroit area and the New York to Boston corridor.

problems for identifying a quantitative correlation between light intensity and population density (particularly in areas of high population density). It also presents difficulties for transforming either variable to improve the correlation. This correlation could perhaps be improved if the percent detection dataset or the low-gain dataset were used.

The population density image has a distinctly different distribution (Figure 2a). Like the DMSP OLS image, the most common values are the low values of zero and one person per square km (41 percent and 57 percent of the pixels, respectively); however, the frequency of higher values decreases monotonically in a manner suggesting an exponential decay. The values range from a minimum of zero to a maximum of over 50,000 persons/km$^2$.

Another manipulation of the data was performed in which all of the saturated pixels in the DMSP OLS stable-lights image were grouped in such a way that all adjacent saturated pixels were clustered into independent urban "clusters" (Plate 1 with insets). These urban clusters were overlaid over the population density dataset to produce a set of over 5000 paired data points in which the first value was the area of saturation or urban cluster and the second value was the actual population that lived within that area.

These manipulations of the data allow for comparisons of the two datasets in several ways: (1) correlation on a pixel-by-pixel basis, (2) correlation at a variation of scale of resolution on a pixel-by-pixel basis, (3) correlation of aggregated DMSP pixels at state and county level to population of State and County populations, (4) utilization of the spatial nature of the data to cluster saturated "urban" areas and compare them to their corresponding populations. The results of these comparisons follow.

## Results

Table 1 describes the distribution of the values of the overlaid pixels of the two datasets. The first column gives the values for the DMSP OLS pixels; the second column describes their relative frequency. The remaining columns provide descriptive statistics regarding the values of the population density pixels that overlap with the DMSP OLS pixels of that value. The first and last line in this table are the most interesting. The first record or line, for those square kilometres that register a one on the DMSP sensor (these pixels were really zeroes but one was added to all pixels that were on the land to distinguish them from zero pixels on the ocean), constitutes almost 90 percent of the continental United States and only 17 percent of the human population. The remaining 10 percent of the pixels (those with non-zero DMSP values) coincide with over 80 percent of the human population. This may be one of the most valuable pieces of information

TABLE 1. TABLE OF DMSP PIXEL VALUES AND CORRESPONDING POPULATION DENSITY VALUES

| DMSP Pixel Value | # Of Pixels at this value | Mean of Pop Pxls at value | Max of Pop pxls at value | Range of Pop pxls at value | Sum of Pop pxls at this DMSP pxl value | Std Dv Pop pxls | Median pop pxls | % of Total Pop | % of DMSP pxls |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 6880018 | 5.978 | 4258 | 4258 | 41125756 | 18.4 | 1 | 17.42 | 89.34 |
| 12 | 9 | 3.556 | 9 | 9 | 32 | 3.2 | 4 | 0 | 0 |
| 13 | 26 | 4.962 | 22 | 22 | 129 | 5.3 | 6 | 0 | 0 |
| 14 | 63 | 5.159 | 40 | 40 | 325 | 9.3 | 0 | 0 | 0 |
| 15 | 164 | 3.354 | 77 | 77 | 550 | 7.8 | 0 | 0 | 0 |
| 16 | 230 | 5.896 | 76 | 76 | 1356 | 12.9 | 1 | 0 | 0 |
| 17 | 337 | 4.329 | 83 | 83 | 1459 | 9.4 | 1 | 0 | 0 |
| 18 | 441 | 5.739 | 107 | 107 | 2531 | 14.4 | 1 | 0 | 0.01 |
| 19 | 637 | 8.777 | 176 | 176 | 5591 | 20.9 | 2 | 0 | 0.01 |
| 20 | 804 | 9.649 | 327 | 327 | 7758 | 23.3 | 2 | 0 | 0.01 |
| 21 | 1063 | 7.272 | 173 | 173 | 7730 | 14.9 | 1 | 0 | 0.01 |
| 22 | 1283 | 9.44 | 223 | 223 | 12112 | 18 | 3 | 0.01 | 0.02 |
| 23 | 1502 | 9.792 | 278 | 278 | 14708 | 21.4 | 3 | 0.01 | 0.02 |
| 24 | 1855 | 11.57 | 333 | 333 | 21462 | 24.3 | 3 | 0.01 | 0.02 |
| 25 | 2035 | 11.956 | 289 | 289 | 24331 | 24.3 | 3 | 0.01 | 0.03 |
| 26 | 2152 | 13.633 | 393 | 393 | 29338 | 29 | 5 | 0.01 | 0.03 |
| 27 | 2349 | 12.875 | 384 | 384 | 30244 | 26.3 | 4 | 0.01 | 0.03 |
| 28 | 2587 | 13.351 | 1124 | 1124 | 34538 | 30.6 | 5 | 0.01 | 0.03 |
| 29 | 2936 | 16.147 | 657 | 657 | 47407 | 35.1 | 5 | 0.02 | 0.04 |
| 30 | 3128 | 15.004 | 1332 | 1332 | 46932 | 36.6 | 5 | 0.02 | 0.04 |
| 31 | 3194 | 16.413 | 581 | 581 | 52423 | 30.1 | 6 | 0.02 | 0.04 |
| 32 | 3494 | 17.082 | 471 | 471 | 59685 | 31 | 6 | 0.03 | 0.05 |
| 33 | 3555 | 15.82 | 512 | 512 | 56239 | 28.2 | 6 | 0.02 | 0.05 |
| 34 | 3883 | 18.218 | 666 | 666 | 70740 | 39.4 | 6 | 0.03 | 0.05 |
| 35 | 4017 | 18.903 | 2014 | 2014 | 75934 | 46.7 | 7 | 0.03 | 0.05 |
| 36 | 4165 | 19.521 | 1892 | 1892 | 81306 | 47.8 | 7 | 0.03 | 0.05 |
| 37 | 4386 | 18.119 | 643 | 643 | 79472 | 33.2 | 7 | 0.03 | 0.06 |
| 38 | 4683 | 20.99 | 916 | 916 | 98295 | 40 | 8 | 0.04 | 0.06 |
| 39 | 4919 | 21.364 | 821 | 821 | 105088 | 40.3 | 7 | 0.04 | 0.06 |
| 40 | 5143 | 23.085 | 738 | 738 | 118726 | 44.3 | 9 | 0.05 | 0.07 |
| 41 | 5264 | 21.343 | 1111 | 1111 | 112347 | 43.6 | 9 | 0.05 | 0.07 |
| 42 | 5353 | 22.644 | 721 | 721 | 121211 | 39.5 | 10 | 0.05 | 0.07 |
| 43 | 5380 | 23.022 | 721 | 721 | 123859 | 40.6 | 10 | 0.05 | 0.07 |
| 44 | 5772 | 25.673 | 702 | 702 | 148186 | 45.6 | 11 | 0.06 | 0.07 |
| 45 | 5862 | 25.413 | 1157 | 1157 | 148970 | 43.9 | 12 | 0.06 | 0.08 |
| 46 | 6264 | 24.107 | 893 | 893 | 151004 | 42.3 | 10 | 0.06 | 0.08 |
| 47 | 6434 | 25.347 | 1333 | 1333 | 163085 | 44.9 | 11 | 0.07 | 0.08 |
| 48 | 6685 | 29.073 | 1940 | 1940 | 194354 | 56.1 | 13 | 0.08 | 0.09 |
| 49 | 6848 | 28.394 | 979 | 979 | 194443 | 49.8 | 12 | 0.08 | 0.09 |
| 50 | 7154 | 29.488 | 985 | 985 | 210955 | 50.2 | 14 | 0.09 | 0.09 |
| 51 | 7110 | 30.611 | 1418 | 1418 | 217643 | 51.9 | 14 | 0.09 | 0.09 |
| 52 | 7505 | 30.46 | 1079 | 1079 | 228599 | 55.7 | 14 | 0.1 | 0.1 |
| 53 | 7649 | 31.493 | 1469 | 1469 | 240890 | 54.1 | 15 | 0.1 | 0.1 |
| 54 | 7501 | 32.443 | 1440 | 1440 | 243354 | 55.4 | 15 | 0.1 | 0.1 |
| 55 | 7674 | 32.101 | 1282 | 1282 | 246343 | 53.8 | 15 | 0.1 | 0.1 |
| 56 | 8070 | 34.076 | 965 | 965 | 274991 | 60.4 | 15 | 0.12 | 0.1 |
| 57 | 8418 | 33.211 | 2009 | 2009 | 279571 | 62.9 | 16 | 0.12 | 0.11 |
| 58 | 8376 | 36.364 | 3631 | 3631 | 304584 | 78.6 | 18 | 0.13 | 0.11 |
| 59 | 8502 | 36.936 | 1197 | 1197 | 314029 | 61.8 | 18 | 0.13 | 0.11 |
| 60 | 9331 | 37.82 | 2087 | 2087 | 352903 | 62.5 | 20 | 0.15 | 0.12 |
| 61 | 9236 | 37.699 | 1247 | 1247 | 348186 | 59.8 | 20 | 0.15 | 0.12 |
| 62 | 9501 | 40.73 | 2000 | 2000 | 386975 | 73.8 | 19 | 0.16 | 0.12 |
| 63 | 9597 | 41.341 | 2220 | 2220 | 396746 | 72.6 | 20 | 0.17 | 0.12 |
| 64 | 586561 | 321.36 | 53465 | 53465 | 188497792 | 798.3 | 7 | 79.83 | 7.62 |

with respect to how the DMSP OLS data can act as a proxy for population parameters. The DMSP OLS imagery locates over 80 percent of the continental United States population on only 10 percent of the land. The third column of the table is basically a measure of the average population density for pixels at the corresponding value of the DMSP OLS pixel. There is a clear trend towards increasing population density with increasing DN values for the DMSP OLS imagery; however, this trend makes a dramatic jump at the last DMSP value (i.e., from 63 to 64). The mean population density value increases from around ten persons per square kilometre at a DN of 1 to 41 persons per kilometre at a DN of 63, but it jumps to 321 at the next increment of the DMSP value. This is a result of the saturation of the DMSP sensor in what are primarily heavily urbanized areas. This high incidence of saturation is clearly a result of choosing the dataset for which the pixel value is the maximum observed value rather than the percentage of times a signal was seen. Further exploration with both the low-gain DMSP OLS dataset and the percent detection version may show stronger direct relationships with some population parameters on a pixel-by-pixel basis.

The fact that the DMSP image is saturated in heavily populated urban areas suggests that aggregation of both images to lower resolution may improve the correlation. Such a manipulation could "buffer" out the effects of saturation and

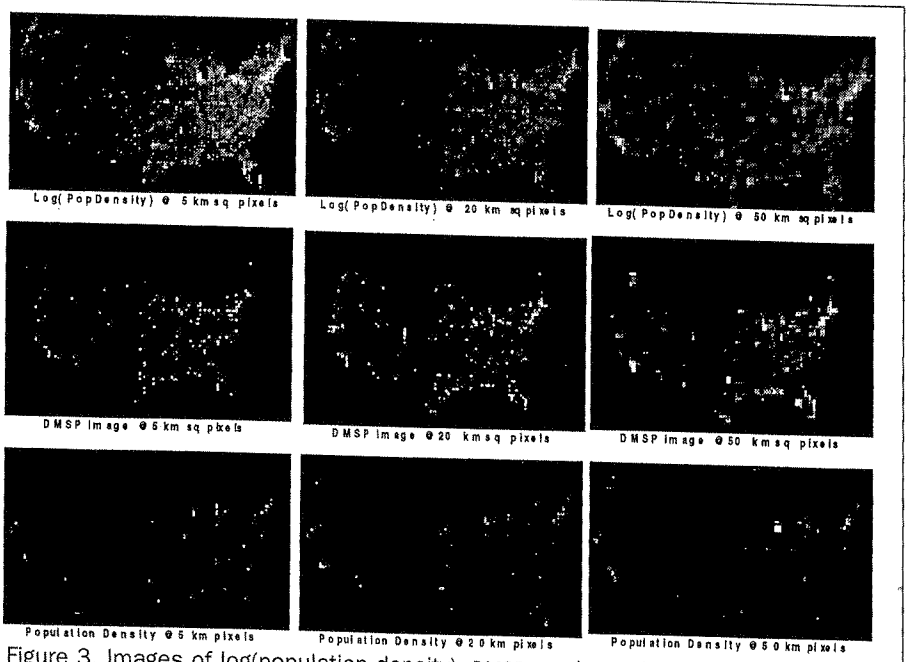



Figure 3. Images of log(population density), DMSP, and population density aggregated to resolutions of 5-, 20-, and 50-square-kilometre pixels.

mitigate any influences of mis-registration. Figure 3 shows images of the DMSP data, population density, and the log of population density at 5-, 20-, and 50-square-kilometre resolutions.

Data on the percentage of the population of each state that lived in urban areas were also obtained. Thus, by multiplying total population by percent urban, as defined by the Bureau of the Census, it was possible to run a regression between the urban population of each state versus the integrated value of the DMSP imagery for each state. This was done to test the hypothesis that the DMSP imagery was really better at being a proxy measurement of the urban population of states. This regression actually resulted in a slightly lower $R^2$ than the one that simply compared total population and the integration of the DMSP pixels in the state.

The first and most direct comparison of these two images was a simple correlation between the raw value of population density pixel and the DMSP pixel value. This cross-correlation resulted in a correlation coefficient of $r = 0.26$. One explanation for the low correlation is simply the fact that the DMSP saturates in most of the areas of high population density. The distribution of the DN values of the two images clearly shows why direct correlations will be problematic. Transforming the population density data by taking its natural logarithm is one means of improving the correlation. The correlation between the DMSP image and the log (population density) raises the correlation coefficient to $r = 0.40$. This result remains unsatisfactory and there is no obvious physical justification for performing such a transformation.

Another transformation involves making use of the spatially referenced nature of the data. Aggregating the pixels to a coarser resolution and using the mean of the constituent pixels as the new larger pixel value usually results in an improved correlation. If there is a scale at which a dramatic change in maximum correlation occurs, it may suggest an appropriate scale for performing analyses of this nature. Aggregation for both the population density versus DMSP image and the log of the population density versus the DMSP image improved correlation as a function of scale (Figures 4a and 4b). Correlation does increase as resolution decreases; however, it does not increase dramatically at any particular

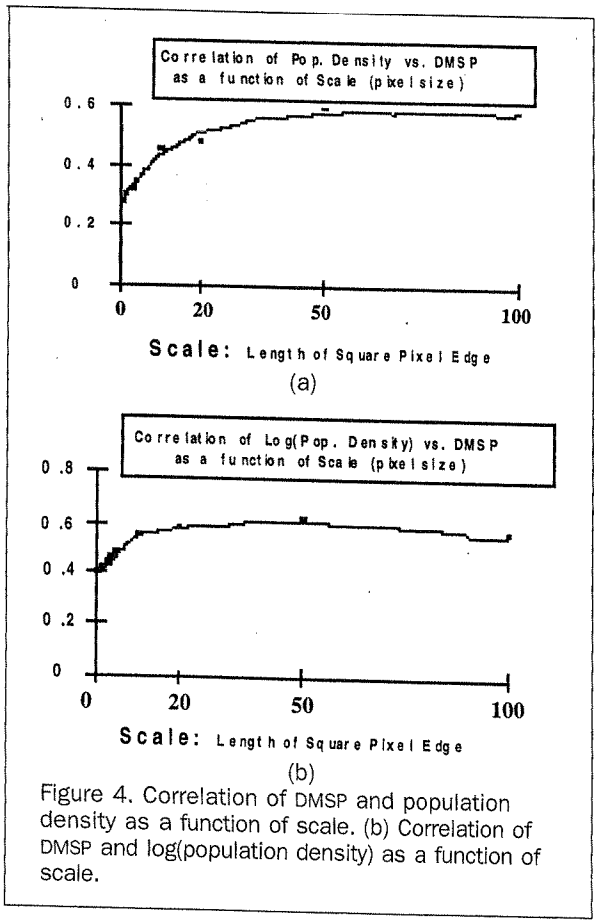



Figure 4. Correlation of DMSP and population density as a function of scale. (b) Correlation of DMSP and log(population density) as a function of scale.

point. The correlation does not maximize at any point between 1 km² and 100 km² for the Population Density versus DMSP images, yet it does reach a maximum somewhere between 20 km² and 100 km² for the log (Population Density) versus DMSP images. Nonetheless, the changes in correlation are not dramatic enough to suggest that there is an optimum scale of resolution for performing these analyses. Correlations generally increase as spatial resolutions decrease for most analyses of this nature.

Aggregating DMSP light values to state and county levels of aggregation is another means of investigating the correlation between population density and DMSP light value. A vector image of the 48 continental United States was used to generate 49 datapoints (each of the lower 48 plus the District of Columbia), in which the predictor value is an integration or sum of all the DMSP pixel values that existed in each state and the predicted value is the population of the state. A regression on these 49 points produced an $R^2$ of 0.69. A similar analysis of all the counties of the United States produced an $R^2$ of 0.50. These regressions were also run on the natural log of the population of the states and counties against the integrated DMSP pixel values, respectively, resulting in $R^2$ of 0.61 and 0.48, respectively (Figures 5a through 5d).

It does not appear that there is any strong quantitative relationship between the intensity of light emitted from a particular place in the United States as measured by the DMSP sensor and the population density of that place. Most of the lack of correlation is undoubtedly due to the saturation of the DMSP sensor. A reclassification of the DMSP image in which all pixels with a value of 63 were set to NODATA resulted in a correlation of $r = 0.02$. Clearly, the intermediate DMSP values from 2 to 63 do not contribute to the correlation to any great extent. These results do suggest that an increased dynamic range of the DMSP OLS instrument could improve the correlation. The striking qualitative correlation between the two images suggests that these methods are not capturing an important facet of the relationship between population density and the nighttime light emissions. The spatial nature of the DMSP imagery is one of its strongest attributes with respect to utilizing it as a proxy for population. Unfortunately, the DMSP OLS stable-lights data do not show a strong correlation with population density within saturated urban clusters. However, by utilizing the spatial nature of the data, the theory of exponential decay of population density from the center to the edges of urban areas suggests a method for estimating the total population of the whole urban cluster from it's area alone. We were interested in confirming this theory of exponential decay of population density with a visual representation of the errors from a simple linear regression analysis.

A simple linear regression model was developed to use the value of the DMSP pixel to predict the population density of the matching pixel. This was done on the 1-km² resolution data. The F-test was significant to the 0.999 level and the $R^2$ was the 0.26 that was previously mentioned. This spatial analysis of residuals proved to be quite interesting. The model was built as follows:

Population Density = Constant + Coefficient×(DMSP DN value), resulting in the following parameters:

Intercept = 9.5; Slope = 3.3.

An image of the predicted value of the population density based on these regression coefficients was produced and subtracted from the actual population density image. This image was reclassed into categories in which overestimates are depicted in green and underestimates in red. The image is black in areas where the estimate was close to the population density. Detailed insets for this image are produced in Plate 2 for selected urban areas in the United States. These insets clearly show a non-random pattern in the errors associated with the prediction of population density from DMSP imagery. DMSP imagery underestimates the population density of urban centers and overestimates the population density of suburban areas. And as stated before, it is fairly accurate at identifying places of low population density. It is also likely that the green areas indicate areas of suburban population growth. These results are well in keeping with the theory of exponential decay of population density from the centers of urban areas.

The images of the residuals of the linear regression clearly suggest that there may be another means of identifying a strong correlation between population density and nighttime light emission. The best means of extracting this strong quantitative measure of this correlation proved to be a manner virtually identical to the previously mentioned methods of Tobler (1969), Welch (1980), Boyce (1963), Maher and Bourne (1969), Nordbeck (1965), and Stewart and Warntz (1958).

The means by which this was accomplished was alluded to earlier in the description of the methods. The DMSP image was grouped by its adjacent saturated pixels into what are presumed to be urban clusters. Plate 1 is an image of some of these urban clusters. The largest three urban clusters by area were the New York metropolitan area (which included almost all of Long Island and extended parts of New Jersey and Connecticut), the Chicago metropolitan area, and the Los Angeles basin. The areas of these urban clusters (as defined by the number of saturated pixels from the DMSP image) were plotted against the population that lived within these urban clusters as derived from the grid of the U.S. census data. A linear regression and an exponential model were fit to these points and resulted in $R^2$ values of 0.84 and 0.93, respectively. The influential nature of the extreme points such as the New York, Chicago, and Los Angeles areas suggest that a transformation of the data is called for. The following theoretical model is given:

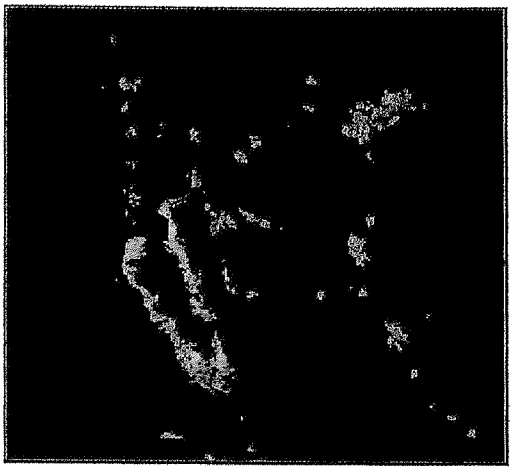
$$R = (a)\times(\text{Pop})^{(b)}.$$

Taking the natural logarithm of both sides of this relationship results in a linear relationship between log(Radius) and log(population). The intercept of this line is the "$a$" parameter and the slope of this line is the "$b$" parameter. Figure 6 is a plot of the natural log of the area of the urban clusters versus the natural log of the population within those urban clusters. A linear regression between these two variables produced an $R^2$ of 0.62. This is somewhat lower than the others because the influence of the large values is dramatically mitigated by this transformation of the data. A quick glance at this plot might suggest that there is a problem with heteroskedasticity or unequal variance. This appearance may merely be a result of the paucity of data points at higher values or it may indicate that the relationship is even stronger for larger urban areas (i.e., has lower variance for large values). This analysis was made quite simple with the use of a GIS. It allowed for the identification of a strong correlation between light emission and population density by taking advantage of the spatial information inherent in the data.

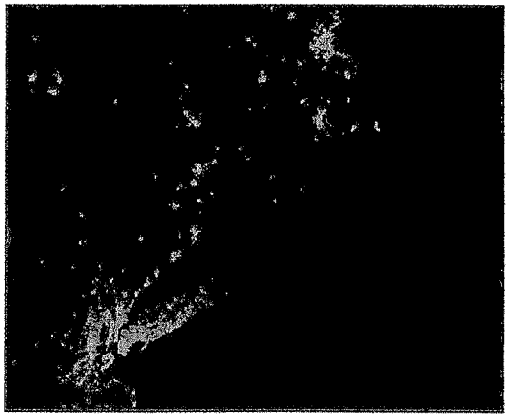
## Conclusion

Despite the fact that DMSP imagery is the most sensitive satellite data available for monitoring VNIR nighttime emissions, it does not show a strong simple quantitative correlation with human population density. This correlation may be greatly improved if the dynamic range and/or the spectral or spatial resolution of the sensor were improved. However, the DMSP imagery serves useful purposes in other ways. Clearly, it is an indicator of human presence in a powerful qualitative way in the United States. Saturated pixels capture over

Seattle and Puget Sound Area



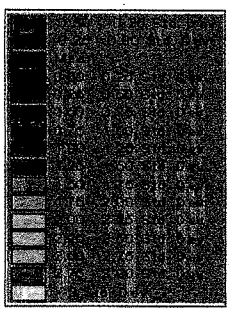
San Francisco Bay Area



New York City, Long Island, Boston, & Cape Cod

Dallas and Fort Worth Area

The Florida Peninsula



New Orleans and Mississippi Delta

Plate 2. Insets of error images for six regions of the United States. Values are actual population density minus regression estimate. Red areas are overestimates, green areas are underestimates, and black areas are accurate. The distribution of these residuals is not normal (see legend). Areas outside saturated urban clusters are in black. The standard deviation of residuals is 814 persons per square kilometre.

80 percent of the population on only 10 percent of the land. This alone is a powerful indication of the spatial distribution of the human population in the United States. This spatial relationship can be augmented in a quantitative way by taking advantage of its spatial information. Saturated areas of imagery can be grouped into clusters. The area of these clusters shows a strong correlation with the population of the area covered. Earlier research suggests that the parameters that influence this relationship will vary from one region of the world to the next. If these parameters can be either identified for other parts of the world or shown to be related to simple national aggregate statistics such as percent of population living in urban areas, GDP per capita, and/or energy consumption, the methods described could prove to be very useful. If systematic relationships between the parameters described here and aggregate national figures can be identified, this method could be used in other parts of the world where good spatially referenced census data are un-

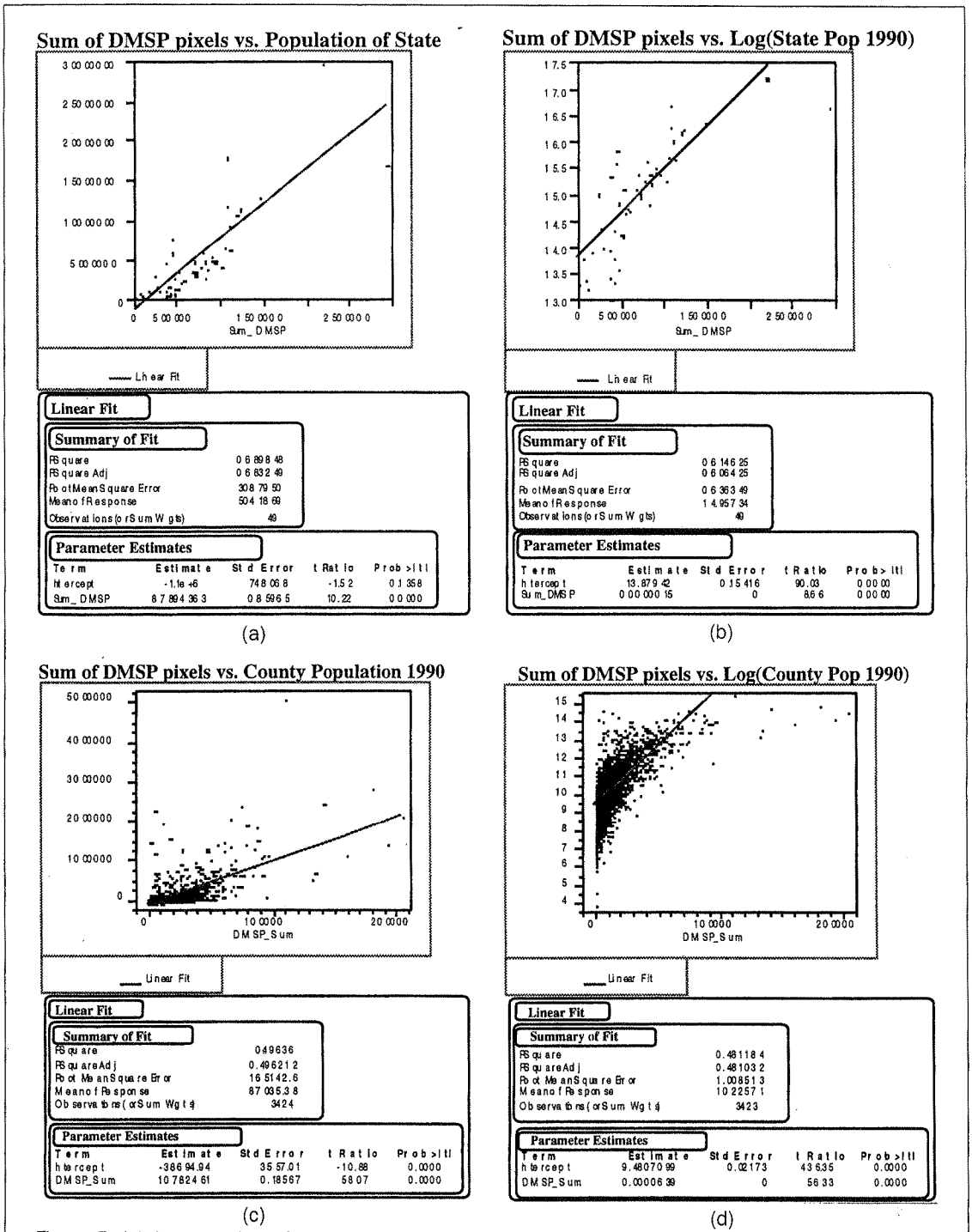



Figure 5. (a) A regression of the integration of all the DMSP pixels in each state versus the population of each state. (b) A regression of the integration of all the DMSP pixels in each state versus the natural log of the population of each state. (c) A regression of the integration of all the DMSP pixels in each county versus the population of each county. (d) A regression of the integration of all the DMSP pixels in each county versus the natural log of the population of each county.

available. Then the DMSP imagery as it presently exists could be incorporated into producing improved maps and datasets of the distribution of the human population in many parts of the world where data of this nature are unavailable. The DMSP imagery might also be used as a primary informant to a smart interpolation program for modeling human population distributions in areas where only large scale aggregate data are available. Present candidates for use as informants to "smart interpolation" programs are city locations, coastlines, topography, railroads, airports, harbors, and rivers. DMSP imagery clearly identifies most if not all major cities, populated coastlines, and some rivers (a DMSP image of Egypt clearly shows the twists and turns of the Nile river).

Further research utilizing DMSP imagery over areas with

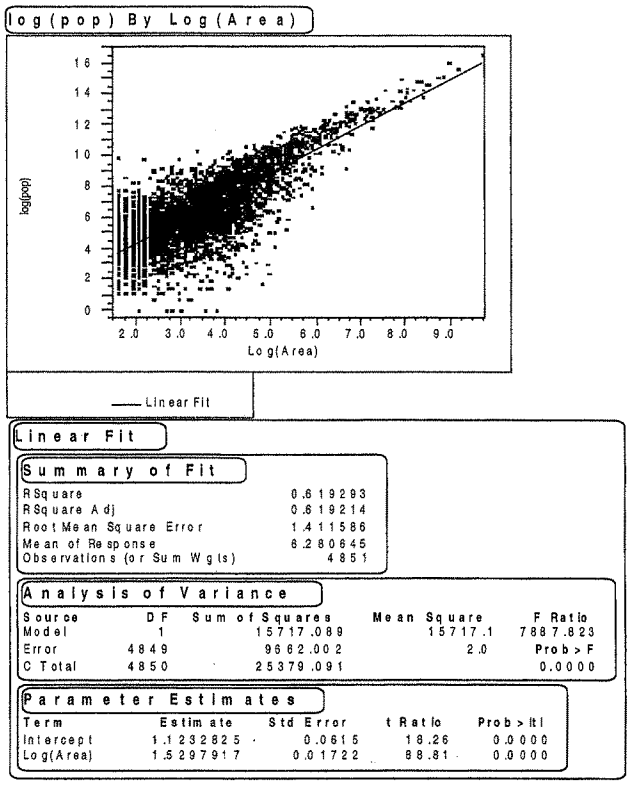



Figure 6. Log(area of urban cluster) versus Log(population of cluster) for all the urban clusters identified from DMSP imagery saturation.

varying degrees of economic development may provide valuable insights into identifying how these parameters vary. Perhaps a systematic relationship can be identified between the area of saturated DMSP pixels over cities, the population of the cities, and some readily available aggregate national statistics such as GDP per capita, percent of population in urban areas, per capita energy consumption, and/or characterizations of the distribution of wealth in the counties in question. If such a relationship could be identified, then the DMSP imagery could prove to be a powerful means of measuring/modeling/monitoring the distribution of the human population at a global scale.

## Acknowledgments

This research was supported in part by the National Center for Geographic Information and Analysis (NCGIA), the Department of Geography at the University of California at Santa Barbara, and a NASA Earth System Science Fellowship. Additional support was provided by the Environmental Systems Research Institute (ESRI) for the development of an educational module for the utilization of Geographic Information Systems (GIS) in secondary education. An education module entitled "Light Up Your Nation" that contains much of the data used for this research and other related information will be available as a technical report from the NCGIA in the spring of 1997.

## References

Boyce, R., 1963. Changing Patterns of Urban Land Consumption, *Professional Geographer*, 15:19–24.

Clark, J., and D. Rhind, 1992. *Population Data and Global Environmental Changes*, International Social Science Council, Programme on Human Dimensions of Global Environmental Change, UNESCO, Paris.

Clayton, C., and J. Estes, 1980. Image Analysis as Check on Census Enumeration Accuracy, *Photogrammetric Engineering & Remote Sensing*, 46(6):757–764.

Croft, T.A., 1977. *Nocturnal Images of the Earth from Space* (Publication No. PB-273 661, prepared for the Topographic Division of the U.S. Geological Survey), Stanford Research Institute, Stanford, California, 113 p.

———, 1978. Nighttime Images of the Earth from Space, *Scientific American*, 239(July):86–98.

Deng, L., 1994. China's Net Migration Surplus Analysis, *Association of American Geographers*, San Francisco, California, 82 p.

Ehrlich, P., 1988. The Loss of Diversity: Causes and Consequences, *Biodiversity* (E.O. Wilson, editor) National Academy Press, Washington, D.C., pp. 21–27.

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(6):727–734.

Forster, B.C., 1985. An Examination of Some Problems and Solutions in Monitoring Urban Areas from Satellite Platforms, *International Journal of Remote Sensing*, 6:139–151.

Foster, J.L., 1991. Observations of Snow and Ice Features During the Polar Winter Using Moonlight as a Source of Illumination, *Remote Sensing of Environment*, 37:77–88.

Maher, C., and L. Bourne, 1969. *Land Use Structure and City Size: An Ontario Example*, Center for Urban and Community Studies, University of Toronto, 124 p.

Meij, H., 1995. *Integrated Datasets for the USA*, Socioeconomic Data and Applications Center (SEDAC) (ftp site: ftp.ciesin.org), Consortium for International Earth Science Information Network.

Nordbeck, S., 1965. *The Law of Allometric Growth*, Michigan Inter-University Community of Mathematical Geographers, Paper 7, 40 p.

Ogrosky, C.E., 1975. Population Estimation from Satellite Imagery, *Photogrammetric Engineering & Remote Sensing*, 41:707–712.

Skole, D.L., 1994. Data on Global Land-Cover Change: Acquisition, Assessment, and Analysis, *Changes in Land Use and Land Cover: A Global Perspective* (William B. Meyer and B.L. Turner, editors), Cambridge University Press, pp. 437–472.

Stewart, J., and W. Warntz, 1958. Physics of Population Distribution, *Journal of Regional Science*, 1:99–123.

Tobler, W., 1969. Satellite Confirmation of Settlement Size Coefficients, *Area*, 1:30–34.

Tobler, W.R., U. Deichmann, J. Gottsegen, and K. Malloy, 1995. *The Global Demography Project*, Technical Report No. 95-6, National Center for Geographic Information and Analysis, UCSB, Santa Barbara, California, 75 p.

Tolba, M.K., 1992. Human Settlement, *The World Environment 1992*, Chapman and Hall on behalf of UNEP, pp. 61–104.

Welch, R., 1980. Monitoring Urban Population and Energy Utilization Patterns from Satellite Data, *Remote Sensing of Environment*, 9(1):1–9.

Welch, R., and S. Zupko, 1980. Urbanized Area Energy Utilization Patterns from DMSP Data, *Photogrammetric Engineering & Remote Sensing*, 46(2):201–207.

(Received 06 May 1996; accepted 05 November 1996; revised 10 March 1997)