SPATIAL ANALYSIS OF GLOBAL URBAN EXTENT FROM NIGHT LIGHTS

Christopher Small[1], Francesca Pozzi[2] and C.D. Elvidge[3]

[1]*Lamont Doherty Earth Obs. Columbia University Palisades NY, 10964 USA*

[2]*CIESIN Columbia University Palisades, NY 10964 USA*

[3]*NOAA National Geophysical Data Center 325 Broadway, Boulder, CO 80303 USA*

KEY WORDS: global, urban, night, light, Landsat

ABSTRACT:

Previous studies of DMSP-OLS stable night lights have shown encouraging agreement between temporally stable lighted areas and various definitions of urban extent. However, these studies have also highlighted an inconsistent relationship between the actual lighted area and the boundaries of the urban areas considered. Applying detection frequency thresholds can reduce the spatial overextent of lighted area ("blooming") but thresholding also attenuates large numbers of smaller lights and significantly reduces the information content of the night lights datasets. Spatial analysis of the widely used 1994/95 stable lights data and the newly released 1992/93 and 2000 stable lights datasets quantifies the tradeoff between blooming and attenuation of smaller lights. For the 92/93 and 2000 datasets, a 14% detection threshold significantly reduces blooming around large settlements without attenuating many individual small settlements. The corresponding threshold for the 94/95 dataset is 10%. Spatial analysis of light size distributions reveals that conurbations larger than 80 km diameter account for < 1% of all settlements observed but account for about half the total lighted area worldwide. Comparison of lighted area with built area estimates from Landsat imagery of 17 cities shows that lighted areas are consistently larger than even maximum estimates of built areas for almost all cities in every light dataset. Thresholds >90% can often reconcile lighted area with built area in the 94/95 dataset but there is not one threshold that works for a majority of the 17 cities considered. Comparison of lighted area with blooming extent for 10 lighted islands suggests a linear proportionality of 1.25 of lighted to built diameter and an additive bias of 2.7 km. While more extensive analyses are needed, a linear relationship would be consistent with a physical model for atmospheric scattering combined with a random geolocation error.

INTRODUCTION

Satellite imaging of stable anthropogenic lights provides an accurate, economical and straightforward way to map the global distribution of urban areas. Urban areas account for a small fraction of Earth's surface area but exert a disproportionate influence on their surroundings in terms of mass, energy and resource fluxes. The spatial distribution and size-frequency characteristics of the global urban network have important implications for disciplines ranging from economics (e.g. [*Fujita et al.*, 1999; *Krugman*, 1996] to ecology (e.g. [*Cincotta et al.*, 2000]) to astronomy ([*Cinzano et al.*, 2001]). In spite of its importance, accurate representations of global urban extent are difficult to derive from administrative definitions ([*Balk et al.*, 2004]). While there are many irreconcilable administrative definitions of urban extent currently in use, physical measurements of lighted area can provide a self-consistent metric on which to base comparative analyses of urban extent. Temporally stable upwelling radiance is unique to anthropogenic land use and can be measured by the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) system [*Croft*, 1978]. However, there are caveats inherent to this characterization of urban

extent. Specifically, the area and intensity of illumination is known to vary significantly with energy availability, economic development and density of settlement [*Elvidge et al.*, 1997; *Sutton et al.*, 1997]. Some of these variations have been quantified at national scales but direct comparisons of urban extent and lighted area have only been done for a few cities.

Previous analyses have revealed a consistent disparity between various spatial measures of urban extent and the spatial extent of lighted areas in the night lights datasets [*Welch*, 1980; *Welch and Zupko*, 1980]. Specifically, the lighted areas detected by the OLS are consistently larger than the geographic extents of the settlements they are associated with. The larger spatial extent of lighted area, relative to developed land area, is sometimes referred to as "blooming". The blooming is the result of three primary phenomena, including: 1) the relatively coarse spatial resolution of the OLS sensor and the detection ofdiffuse and scattered light over areas containing no light souce , 2) large overlap in the footprints of adjacent OLS pixels, and the accumulation of geolocation errors in the compositing process (Elvidge et al., 2004). In the context of this study, blooming refers to spurious indication of light in a location that does not contain a light source.

In order to offset the blooming effect, Imhoff et al ([*Imhoff et al.*, 1997]) proposed using a threshold of 89% detection frequency to eliminate less frequently detected lighted pixels at the peripheries of large urban areas. Imposing a detection frequency threshold effectively shrinks the lighted areas so they are more consistent with administrative definitions of urban extent. The drawback associated with detection frequency thresholds is that they also attenuate large numbers of smaller, less frequently detected settlements. These studies all suggest that it may be possible to obtain reasonable agreement between lighted area and various measures of city size but these studies also reveal significant variability in the relationship between lighted area and different definitions of urban extent. All of these studies have emphasized the need for more extensive analyses of lighted area, detection thresholds and urban extent.

The objectives of this analysis are to quantify the global size-frequency distribution of stable contiguous lighted areas and to investigate the correspondence between the spatial extent of urban land use and lighted area. In the context of this study, we refer to all developed non-agricultural land (i.e. urban, suburban, exurban) as "urban". We conduct a series of comparative analyses of the 94/95 light dataset used in previous studies and the new 2000 and 92/93 datasets recently released by Elvidge *et al.* (2004). Because previous studies have highlighted the importance of detection frequency thresholds on lighted area, we first quantify the dependence of lighted area and size frequency distributions on detection frequency threshold. Because the light data provide a unique measure of urban morphology across a wide range of sizes, we also quantify the shape distributions as area/perimeter ratios. In the second part of the analysis we compare lighted areas with Landsat-derived estimates of urban land use for 17 cities worldwide. We also compare the extent of blooming to lighted area for 10 islands of varying size. The objective of these comparisons is to quantify the spatial overextent of lighted area for each dataset for different detection frequency thresholds to determine if there is a consistent relationship that can be used to correct for the spatial overextent. The overall purpose of this analysis is to quantify the systematics of the global distributions of size, shape and frequency of detection for different OLS night lights datasets and to illustrate their correspondence to other physical measures of urban land use. By quantifying the physical consistencies of the lights data we hope to facilitate future analyses of the non-physical (i.e. socioeconomic, cultural, political) determinants of urban extent and stable radiance.

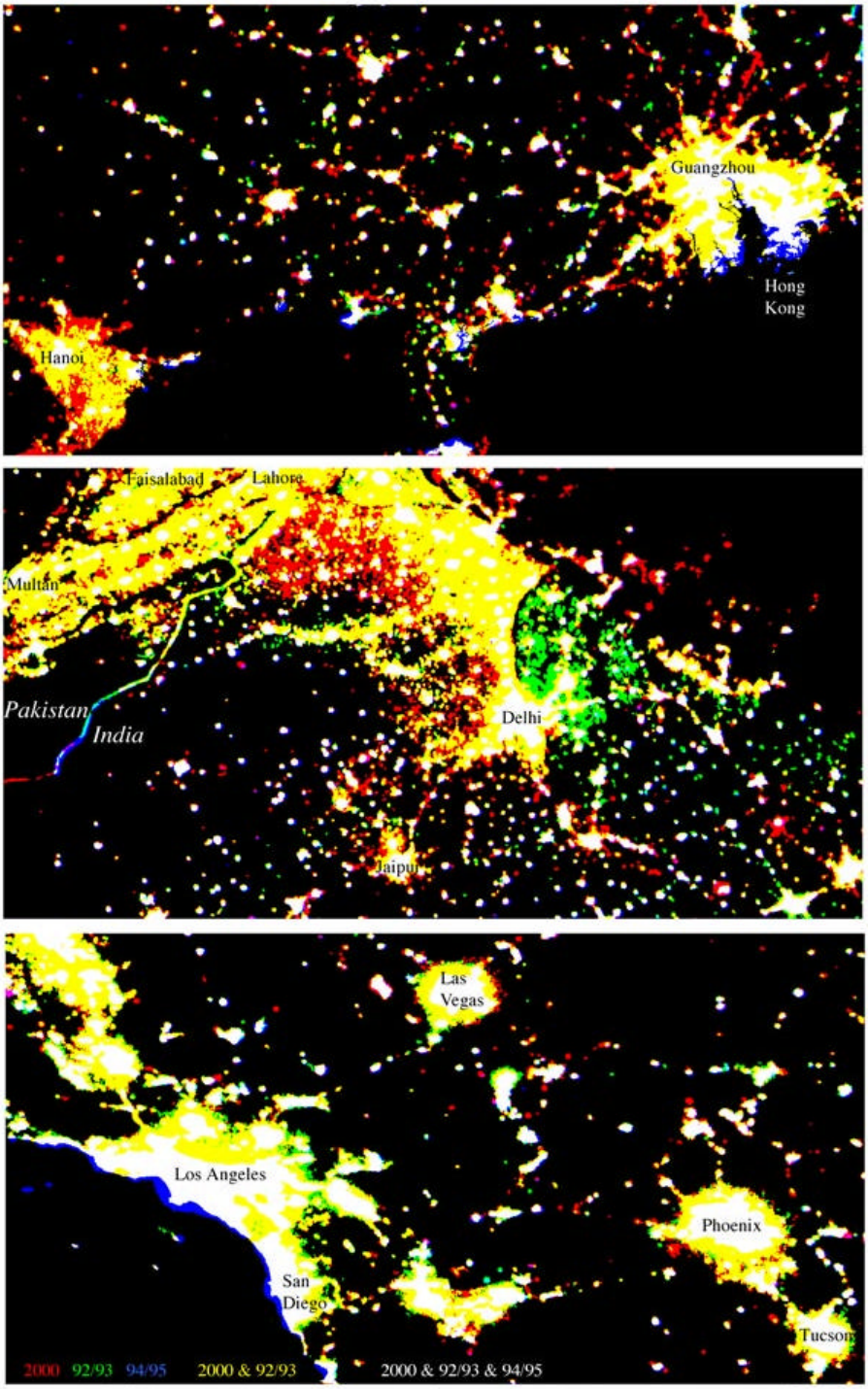

Figure 1. Night light composites. Each dataset is represented by a primary color (R/G/B) while areas lighted in two or three datasets are represented as additive colors (white & yellow). The original 94/95 dataset resolves urban cores and moderate sized settlements while the newer datasets resolve smaller settlements and more diffuse lights at settlement peripheries. The 94/95 dataset is not masked at the coastlines so the effect of "blooming" over water is apparent as blue fringes on coastal cities.

Data

The 1994-95 nighttime lights dataset is a cloud-free composite of OLS data collected between October 1, 1994 and March 31, 1995 under low moon conditions (Elvidge, et al., 2001). The processing involved the manual selection of usable orbital segment, semi-automatic cloud detection, and filtering to detect lights relative to the local background. The basic algorithms have been described in Elvidge et al. (1997). The products used in this analysis are percent frequency of detection images in which values between 0 and 100 indicate the percent frequency of light detection within the set of cloud-free observations over the duration of the observations. The lights were separated into four categories (fires, gas flares, fishing boats, and human settlements) through visual interpretation of the percent frequency image.

The 1992-93 and 2000 nighttime lights datasets are cloud-free composites processed specifically with the objective of change detection. The production was considerably more automated than the methods used for the 94/95 product. Data from two years (1992 and 1993) were used for the first time period due to major temporal gaps present at the start of the archive in mid-1992. The processing included automatic cloud detection and a modified light detection algorithm designed to capture dim lighting. Products include the percent frequency (same as 94-95 product) and the average digital number of the detected lights. Fires were separated from human settlements based on their high brightness levels and low persistence. The products are not radiance calibrated due to the lack of on-board calibration and uncertainty in the gain settings.

For consistency with the widely used 94/95 detection frequency data, we calculated percent frequency of detection for the 92/93 and 2000 datasets. Percent frequency of detection is calculated as 100 times the ratio of the cloud-free light count to the cloud-free coverage count. These three light datasets can be compared directly in the form of a three color composite in which the 2000, 92/93 and 94/95 data are assigned to the red, green and blue channels respectively (Fig. 1). In this format, unlighted areas are black while the primary colors show areas represented in only one of the three datasets. Yellow areas highlight the greater spatial extent of the 2000 and 92/93 datasets into areas not illuminated in the 94/95 dataset. Throughout the analysis we refer to clusters of adjacent lighted pixels as contiguous lighted areas. We assume that spatial distributions of these lighted areas reflect characteristics of a detectable subset of human settlements.

Spatial Analysis

Polygons, Centroids and Threshold Area Estimation

In order to quantify the dependence of lighted area and size frequency distributions on detection frequency threshold, we calculated the number of contiguous light polygons, their area, perimeter, and latitude and longitude of their centroids for each dataset (92/93, 94/95 and 2000) at different thresholds. We calculated the number of light polygons and their area and perimeter for each threshold at 10 % intervals between 0% and 100 % frequency, and at 2% intervals between 0% and 30% frequency. To quantify the correspondence between size and frequency of detection, we calculated the detection frequency at the centroid of each polygon at 10% intervals.Because most people find linear distance more intuitive than area, we generally represent the size of each irregularly shaped contiguous lighted area as equivalent circular diameter defined as $2 \times \mathrm{Sqrt}(\mathrm{Area}/\pi)$.

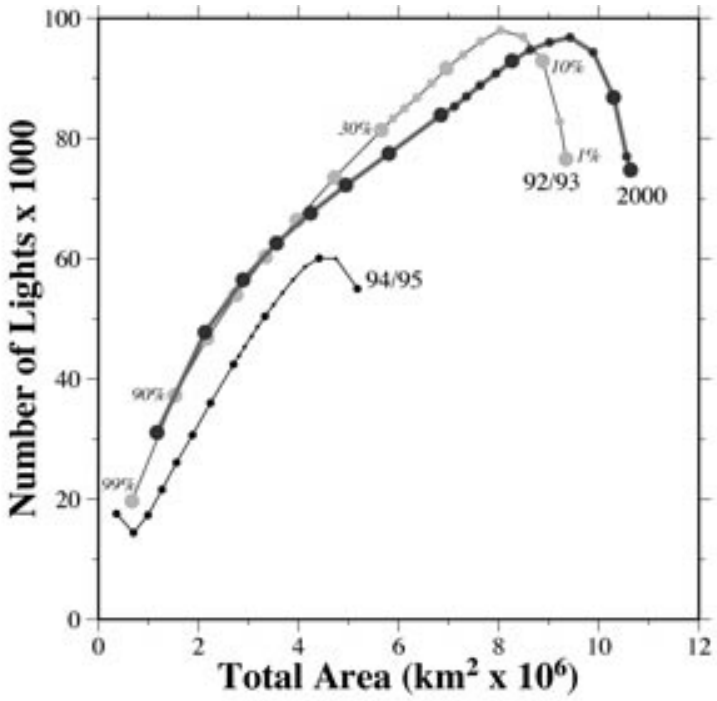

Figure 2. Change in lighted area and number of contiguous lights for different detection frequency thresholds. As thresholds increase from ≥1% to ≥99% the total lighted area diminishes monotonically but the number of contiguous lights initially increases as large conurbations fragment. As thresholds increase further the number of lights diminishes as greater numbers of smaller, less frequently detected, lights are attenuated. A 14% detection frequency threshold results in the maximum number of individual lights for the 2000 and 92/93 datasets while a 10% threshold maximizes the number of lights in the 94/95 dataset. Note that the newer 2000 and 92/93 datasets detect about twice the number of lights and lighted area as the older 94/95 dataset.

Thresholds and Size-Frequency Distributions

Increasing the frequency of detection threshold results in both fragmentation and attenuation of contiguous lighted areas. The changes in the total lighted area and number of contiguous lights with increasing detection threshold shown in Figure 2 illustrate the combined effects of fragmentation and attenuation. Attenuation occurs when smaller, less frequently detected lights fall below the detection frequency threshold and disappear from the threshold-limited map. Increasing the frequency of detection threshold also reduces the lighted area of larger settlements as the lower frequency pixels at the peripheries are exceeded by the threshold value. Fragmentation occurs when a contiguous lighted area subdivides into smaller areas as the detection threshold increases. This corresponds to lighted areas in which two or more centers of high detection frequency are joined into a larger contiguous light by lower frequency pixels in the area between the higher frequency centers. Large conurbations are actually composed of multiple bright, frequently detected urban centers for which peripheral blooming overlaps, plus dim lighting detected outside of traditional urban boundaries. Increasing the detection frequency threshold attenuates the blooming thereby causing the larger contiguous lighted area to fragment into smaller centers of higher detection frequency. Figure 2 shows how the number of contiguous lights initially increases as the frequency detection threshold goes from 0 to 14% (10% for the 94/95 dataset) but then

decreases at higher thresholds. The decrease occurs as the rate of attenuation exceeds the rate of fragmentation.

The size frequency distributions and their variation with % threshold reveal several important properties of the light datasets. The 14% threshold seen in the 92/93 and 2000 datasets may provide a useful cutoff to maximize the number of individual lights thereby balancing attenuation of small individual lights with reduction of spatial overextent around large cities. The size-frequency distributions in Figure3 indicate that increasing the threshold from 10% to 20% results in a sharp increase in the number of very small lights and that this secondary mode of the distrib ution persists as the threshold increases further. The appearance of this secondary mode suggests that much of the area observed at less than the 14% peak threshold is associated with interstitial blooming between settlements rather than individual small lights. The centroid frequency distributions in Figure 4 support this assertion as they show that the distribution of individual lights peaks near the 14% threshold and drops rapidly at lower thresholds. Taken together, these observations suggest that mo st of the lighted pixels below the 14% threshold are associated not with small individual lights but with blooming on the periphery of larger settlements. The areas detected in less than 14% of observations correspond to 14% and 11% of the total areas of the 92/93 and 2000 datasets respectively. Therefore, applying a 14% threshold significantly reduces blooming without attenuating many small settlements. In this sense, the 14% threshold provides the maximum information content for the 92/93 and 2000 datsets – although it still results in significant overestimation of individual city size. The analogous threshold in the 94/95 dataset occurs at 10% detection and corresponds to 15% of the total lighted area.

REFERENCES

Balk, D., F. Pozzi, G. Yetman, A. Nelson, and U. Deichmann, What can we say about urban extents? Methodologies to Improve Global Population Estimates in Urban and Rural Areas?, in *Population Association of America Annual Meeting*, Boston, MA, 2004.

Cincotta, R.P., J. Wisnewski, and R. Engelman, Human population in the biodiversity hotspots, *Nature*, *404*, 990-992, 2000.

Cinzano, P., F. Falchi, C.D. Elvidge, and K.E. Baugh, The artificial sky brightness in Europe derived from DMSP satellite data, in *Preserving the Astronomical Sky*, pp. 95-102, 2001.

Croft, T.A., Nighttime images of the earth from space, *Scientific American*, *239* (86-89), 1978.

Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, E.R. Davis, and C.W. Davis, Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption, *International Journal of Remote Sensing*, *18* (6), 1373-1379, 1997.

Elvidge, C.D., J. Safran, I.L. Nelson, B.T. Tuttle, V.R. Hobson, K.E. Baugh, J.B. Dietz, and E.H. Erwin, Area and position accuracy of DMSP nighttime lights data., in *Remote Sensing and GIS Accuracy Assessment*, pp. 281-292, CRC Press, 2004.

Henderson, M., E.T. Yeh, P. Gong, C. Elvidge, and K. Baugh, Validation of urban boundaries derived from global night-time satellite imagery, *International Journal of Remote Sensing*, *24* (3), 595-609, 2003.

Imhoff, M.L., W.T. Lawrence, D.C. Stutzer, and C.D. Elvidge, A technique for using composite DMSP/OLS "city lights" satellite data to map urban area, *Remote Sensing of Environment*, *61* (3), 361-370, 1997.

Small, C., A global analysis of urban reflectance, in *Proceedings of the Third International Symposium on Remote Sensing of Urban Areas*, edited by D.M.a.F.S.-E. C. Jurgens, Istanbul, Turkey, 2002.

Sutton, P., D. Roberts, C. Elvidge, and K. Baugh, Census from Heaven: An estimate of the global human population using night-time satellite imagery, *International Journal of Remote Sensing*, *22* (16), 3061-3076, 2001.

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