



Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination

Alexei Abrahams^{a,*}, Christopher Oram^c, Nancy Lozano-Gracia^b

^a Niehaus Center, Woodrow Wilson School of Public and International Affairs, Princeton University, United States

^b World Bank, GSURR - Social, Urban, Rural and Resilience

^c TRIUMF, Vancouver, BC, Canada

ARTICLE INFO

Keywords:

DMSP nighttime lights
Blurring
Urban boundaries

ABSTRACT

A well known difficulty with the Defense Meteorological Satellite Program's nighttime lights series (DMSP-NTL 1992–2012) is that the images suffer from pervasive blurring, dubbed ‘overglow’ or ‘blooming’. In this paper we devise a new method that significantly mitigates blurring. We assemble a sample of isolated light sources around the globe and discover that blurring is governed by a symmetric Gaussian point-spread function (PSF), but that the brightness of sources widens the PSF. To make sense of this, we recreate step-by-step the satellite's data collection and storage process, and discover an important fact: any pixel containing a light source will tend to be lit at least as often as its neighbors. This regularity provides a second filter on the data that allows us to calibrate the dimensions of the PSF to each part of the globe, each satellite, and each year. We generate a user-friendly, open-access MATLAB script that deblurs all DMSP-NTL images for all years, and we showcase the enhanced images for a sample of locations around the globe.

1. Introduction

Since at least Croft (1979), we have known that the Defense Meteorological Satellite Program's nighttime lights images (DMSP-NTL) suffer from significant blurring, dubbed ‘blooming’ or ‘overglow’. For each year 1992–2012, the National Oceanic and Atmospheric Administration (NOAA) has undertaken the extraordinary service of compiling these images into annual composite images. Stored as GeoTiff rasters with the identifier “avg_vis.tif”, they are freely accessible to the public (<https://www.ngdc.noaa.gov/eog/dmsp>). The avg_vis images encompass almost all inhabited areas of the globe, recording the average quantity of light observed at each pixel across cloud-free views for each year. The images naturally lend themselves to tracking urban growth and electrification rates of human settlements worldwide, but remote sensing scientists quickly discovered that light from human settlements was spreading far beyond known boundaries. Light from coastal cities was observable 20 km out to sea, and for cities with known administrative boundaries, the lit area would often exaggerate the city's size by a factor of 10 (see Imhoff et al., 1997; Henderson et al., 2003).

Blurring is particularly problematic for social scientific research of urban areas. The immense popularity of DMSP-NTL among social scientists, along with other satellite images, owes to their usefulness as a neighborhood-level proxy for variables like electrification, gross

domestic product (GDP), or population. While governments of developed countries do measure these variables directly, they often fail to do so at a high degree of spatial and temporal disaggregation, and methods of data collection are not standardized across countries. In under-developed or conflict-ridden parts of the world, such data are often not collected at all. DMSP-NTL images by contrast are easily accessible, standardized globally, and available annually for the 1992–2012 period. For these reasons they are often adopted as a proxy. Min (2015), for example, uses DMSP-NTL to show that patronage politics predict communities' electrification rates. Gonzalez-Navarro and Turner (2016) use DMSP-NTL to test if the building of subways predicts subsequent population growth. Baum-Snow et al. (2017) use DMSP-NTL to test if the extension of railroads through Chinese cities has facilitated the decentralization of economic growth. For all of these studies, however, blurring is a confounding factor for neighborhood-level analysis, forcing the researchers to be more tentative about their conclusions or to test their hypotheses at lower resolutions. The sample of affected studies also suffers from attrition, as researchers aware of the blurring problem abandon or even fail to attempt neighborhood-level analyses.

The blurring problem has remained unsolved for two decades. Initially it was noted that each avg_vis image's corresponding pct image, which records the percentage of cloud-free nights on which each pixel was observed to be lit (henceforth *frequency of illumination*), exhibited a

* Corresponding author.

E-mail address: alexei_abrahams@alumni.brown.edu (A. Abrahams).

marked decline in frequency of illumination away from city centers. This led some studies to propose a technique termed ‘thresholding’ wherein all pixels lit below some frequency threshold would be turned off (Imhoff et al., 1997; Henderson et al., 2003). Comparing the results to actual urban boundaries, however, it seemed that the appropriate threshold varied from city to city, and from year to year. To address this, the most recent incarnations of this approach attempt to estimate optimal thresholds for each location-year. Pinkovskiy (2013), for example, uses land cover data to rule out ‘barren’ pixels as non-urban. Zhou et al. (2014) pre-filter DMSP-NTL images with various masks, then trains a logistic learning model on land cover data labeled by MODIS-derived urban classifications. Zhang et al. (2013) avoid hard thresholds, choosing instead to fold together DMSP-NTL and NDVI on the premise that areas with more vegetation are less urban. All of these efforts, while ingenious, suffer from several shortcomings. Firstly, within the social sciences, researchers are interested not only in tracking the urban dynamics of different parts of the world over time, but also explaining them in terms of local features of the urban and peri-urban environment. By absorbing these local features into the urban footprint metric itself, the aforementioned studies render this variable useless for testing many social scientific hypotheses. If, for example, we attempt to improve the accuracy of our urban extent estimates by blending lights and vegetation indices, the resulting composite metric cannot be used to test whether poor (dim) areas of the city enjoy greater access to gardens and parks than rich (bright) areas. On this matter, it should further be noted that while it is broadly correct that NDVI scores and DMSP-NTL values are inversely related across wilderness areas and cities, the inverse relationship *within* cities is more precarious. Indeed, the wealthiest and best lit neighborhoods of cities are also typically the best vegetated. Folding population or land use data (Zhou et al., 2014; Sharma et al., 2016) into the urban estimate likewise precludes any research questions that hope to analyze the relationships between urbanization, electrification, or economic growth on the one hand, and land use or population density on the other. Moreover the ingenuity of these various methods seems to be inversely correlated with their ease of implementation. They draw on a dizzying array of data that may or may not be available globally for the 1992–2013 period. Scenes then have to be mosaicked, fed through a training algorithm or collapsed together by functional forms imposed by the researcher.

In this study we develop a new method to deblur DMSP-NTL images, with several major advantages over previous methods. Firstly, our method is totally independent of auxiliary data like NDVI or land classification; we require as input only the blurred DMSP avg_vis image and its corresponding pct image. For every satellite-year, both of these images are freely available for download from NOAA's website, so users can apply our method to all DMSP-NTL images, 1992–2012 (even the radiance-calibrated images). Since we have uploaded our MATLAB implementation script for free (<https://github.com/alexeiabrahams/nighttime-lights>), the user need only download the requisite image pairs and run our script. Our script runs very fast: with a 2.6 GHz Intel 7th Generation CPU, we find that our MATLAB script deblurs DMSP images at a rate of almost 4000 pixels per second. An urban researcher, for example, can obtain a deblurred rendering of a typical sub-Saharan African capital in 2 to 3 s. By comparison, implementation of NDVI- or land-classification-based methods requires the user to obtain such auxiliary data, mosaic and match scenes to the NTL images, and implement various algorithms to transform the data. Finally, note that all previous efforts to solve the blurring problem have relied on a purely statistical (correlative, machine-learning) approach: while they do mention in passing several possible sources of blurring, they never explain in detail how blurring happens. Instead, they simply introduce an auxiliary data source (urban administrative boundaries, NDVI, classified land type data) and assume some relationship between those data and light emissions (light cannot be emitted outside urban boundaries; light is unlikely to be emitted from dense vegetation; light cannot be

emitted from land classified ‘empty’). By contrast, our method, though partly statistical, also relies heavily on an intimate understanding of how DMSP satellites work. Indeed, an intermediate contribution of our study is to offer readers a step-by-step explanation of how these satellites collected and stored data.

2. Materials & methods

Our method deblurs images using two filters in sequence. The first filter makes an assumption that light was blurred via a symmetric Gaussian point-spread function (PSF). Below, we assemble a selected sample of 47 isolated light sources from around the globe to demonstrate that blurring does indeed appear to follow a symmetric Gaussian pattern. The filter attempts to invert the blurring process in a noise-sensitive manner, using a standard Wiener Deconvolution. The first filter is therefore very conventional, approaching the problem in the usual manner recommended by standard image processing textbooks (for example, see Hansen et al., 2016).

But as we show with our sample of isolated sources, the dimensions of the PSF (governed by the distribution's standard deviation parameter σ) increase with the brightness of the source. So whereas a relatively dim source may exhibit a Gaussian blur of $\sigma = 1.55$ km, brighter sources can exhibit up to $\sigma = 3.0$ km. To make sense of this, we recreate step-by-step the satellite's entire on-board data collection and data storage process, demonstrating how and why blurring is exacerbated by source brightness. From this exercise we discover an important fact that underwrites our second filter: the pixel at which a light source is located will always be a local maximum in the pct image (it need not be a local maximum in the avg_vis image, of course). Stated another way, the pixel where a light source is located will always be at least as frequently lit as all of its immediate 8 neighboring pixels. The contrapositive implication is that pixels, if they are not local maxima in the pct image, *cannot* be the locations of light sources. Our second filter, therefore, is to turn off (set to zero) in the avg_vis image all pixels that are not local maxima in the pct image. But if in the first filter we chose the correct σ , then the effect of this second filter should be very small: there should be very little residual light left to ‘turn off’. This fact allows us to loop back to the blurred image and apply the first filter with different σ s, and compare the sum of residual light each time. The optimal σ^* (for a particular window, satellite, year) is the σ that minimizes the sum of residual light turned off by the second filter.

2.1. Isolated blurs in the data

Although our ultimate objective is to deblur the whole world, including cities, we cannot actually learn much about the shape or pattern of blurring by looking at cities: there we find many light sources in close proximity to each other, so their blurs overlap and appear distorted. Instead, we must scour the globe for isolated, single-pixel light sources. But this task is made difficult by the fact that many isolated sources are rural villages, which are often dimly lit and therefore cast an even dimmer blur which often gets ‘bottom-censored’ (set to zero) onboard the satellite or during NOAA's data processing. Obviously there also does not exist a census or sampling frame of isolated light sources globally, so the sample we assemble is necessarily selected. We try, nevertheless, to introduce a fair degree of spatial and temporal variation. We find 26 light sources in the F15-2000 image and 21 different sources in the F18-2010 image, for a total of 47 sources. The 26 sources in F15-2000 are all oil rigs in the Persian Gulf (13) and North Sea (13). By visual inspection, the remoteness and smallness of oil rigs help generate the most reliable, cleanest blurring patterns, and their bright flares ensure a sufficiently bright source to avoid the aforementioned bottom-censoring problem. By drawing from the Gulf and the North Sea, we also ensure latitudinal variation. Fig. 1 depicts Persian Gulf object 7 and North Sea object 4. At first glance, the blur cast by the North Sea object appears much more elliptical than the Persian Gulf

Download English Version:

<https://daneshyari.com/en/article/8866599>

Download Persian Version:

<https://daneshyari.com/article/8866599>

[Daneshyari.com](https://daneshyari.com)