



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

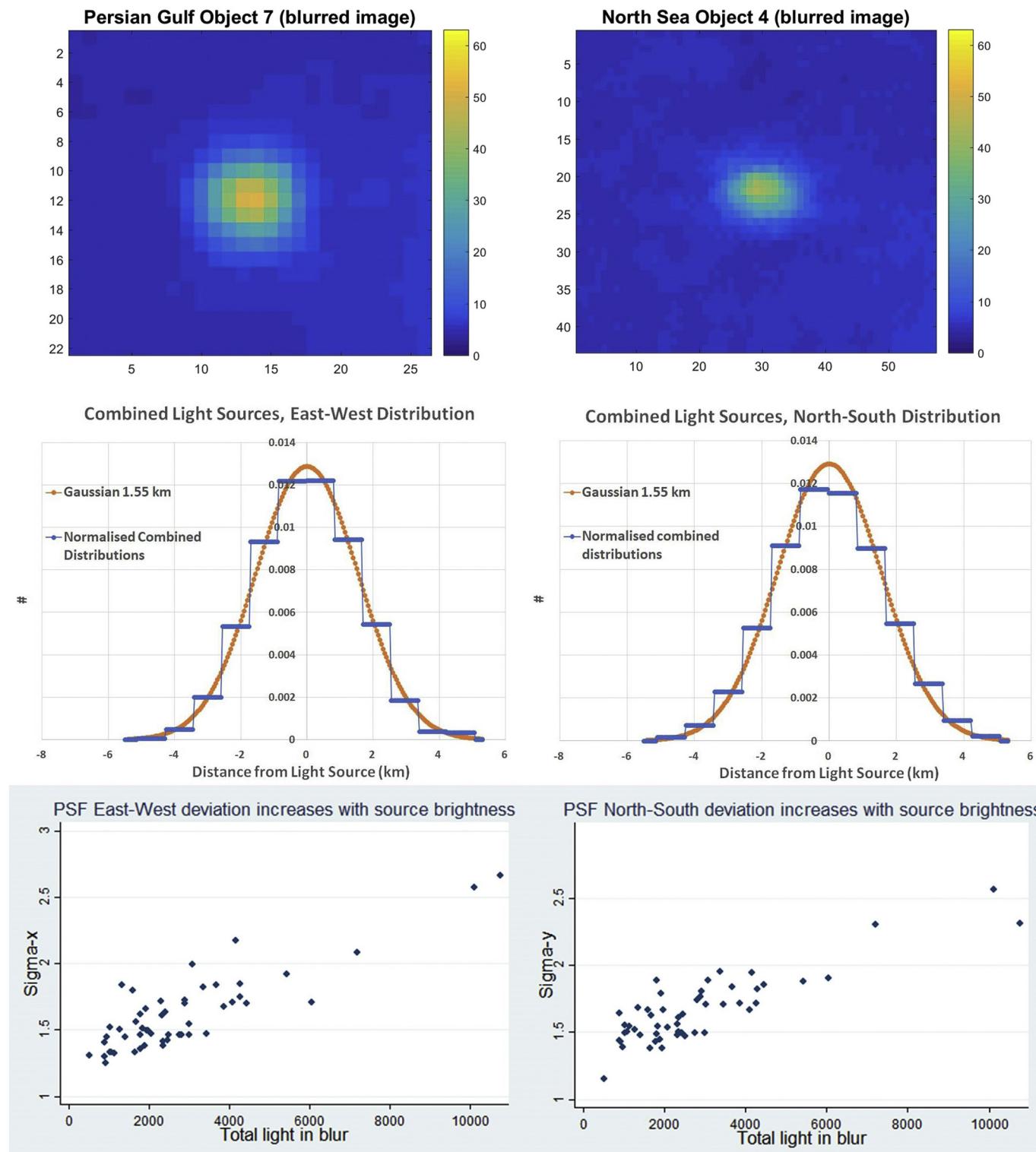


Fig. 1. Isolated lights data analysis.

object, stretched much farther East-West than North-South. This is an illusion, however: both the avg_vis and pct images are displayed in Lambertian coordinates, so whereas pixels throughout the globe have an equal North-South length of approximately 926 m, pixels further from the equator are *narrower* East-West. For example, the width of pixels in the Persian Gulf is over 700 m, whereas pixels in the North Sea have a width closer to 500 m.

The other 21 sources identified in F18-2010 include oil rigs, natural gas flares, and isolated man-made outposts in China, central Asia,

Africa, and the Americas. These 21 sources were chosen also because they generated blurs whose total sum of light tended to be similar, averaging 662 counts. For each light source, we form East-West and North-South histograms by summing pixel values by column and by row, respectively. Taking the average across these 21 sources, we obtain the histograms in Fig. 1, which appear to fit the shape of a symmetric Gaussian distribution with $\sigma = 1.55$ km. The Gaussian shape is consistent with the findings of Small et al. (2005).

If all light sources around the world, no matter how bright,

generated roughly this same Gaussian blur of dimension $\sigma = 1.55$ km, then we could conclude that we have a *spatially invariant PSF*, in which case we could deblur all DMSP-NTL images with a single transformation (Hansen et al., 2016). But fitting a Gaussian distribution to the East–West and North–South histograms of each of our 47 isolated blurs, for each of the satellite-years in which they are viewed, we obtain the graphs in Fig. 1. It is clear from these graphs that blurs generate estimates of σ_x and σ_y that increase with the brightness of the isolated source. In other words, brighter sources cast larger blurs, so our PSF is not invariant and so a one-size-fits-all approach to deblurring will not work. But can we specify the appropriate PSF size by the source's brightness? Unfortunately not: at this point we can only say that the positive relationship between brightness and wideness of the PSF observed in Fig. 1 is correlative: we cannot be sure that there is a *causal* relationship between brightness and PSF size, so we cannot prescribe an appropriate rule to let the PSF dimensions vary with brightness of source.

To make headway, we need to understand the causal mechanisms underlying the correlations observed in Fig. 1. We do this by recreating and simulating step-by-step the entire data collection and data management process of DMSP satellites. It will turn out that first 8-bit and then 6-bit quantization of nightly images compress pixel values in a way that widens the blurs as the sources brighten.

2.2. DMSP-OLS data collection and management

DMSP satellites have been collecting nighttime lights data since the early 1970s, though only since 1992 have these data been organized into global, annual composites easily accessible by researchers. The purpose of these satellites is to monitor weather patterns by spotting cloud formations hovering above the earth's surface (Croft, 1979; Doll, 2008). To accomplish this task, the satellites are equipped with several sensors, including an Operational Linescan System (OLS hereafter) designed to detect light reflecting off the tops of moonlit clouds at night. However, on nights when cloud cover is thin or absent over a given locality, the scanner can see through to the Earth's surface and record nighttime light emissions.

2.2.1. First source of blurring: field-of-view variation

While traveling from pole to pole at an altitude of 833 km, a DMSP satellite's scanner gazes downwards to nadir, then swings its gaze eastward and westward in a 3000-km wide sweep. In Fig. 2, the scanner's instantaneous field of view (ground scanning area, or GSA) is depicted by a red ellipse. At nadir, where the distance from the satellite to the Earth is minimal, the GSA is circular (in actual fact the scanning area is a rotating square (see, for example, Lieske, 1981), but since the angle of rotation at the instant of viewing is uniformly distributed across all nights of the year, the 'average' shape is circular, hence the depiction here). As the scanner swings east or west, it views locations further away, so the circular GSA expands and morphs into an ellipse (longer East–West than North–South). Now assume there is a light source on the Earth's surface, and the satellite's scanner sweeps across it with an elliptical GSA. A quantity of photons rises from the source, and some fraction of that quantity enters the scanner. Photons strike the photocathode, causing a stream of electrons to be released up the photomultiplier tube, amplified by a dynode chain into an electron current. This current is measured and digitized by an analogue-to-digital converter (ADC), thus quantifying the original intensity of light viewed by the scanner. That digital measure of intensity is ascribed to the green, .56 km × .56 km 'fine' pixel centered on the centroid of the elliptical GSA (see the top image of Fig. 2). The scanner then continues its East–West sweep, ascribing the next light intensity to the next fine pixel, and so on. A fixed light source will therefore fall within the boundaries of multiple overlapping elliptical GSAs, so that not only will the light source's own pixel be illuminated, but indeed an entire belt of pixels east and west of the light source will be illuminated, stretching

the full major diameter of the elliptical GSA centered at the light source (see the bottom-left image of Fig. 2). Since the satellite is also advancing North–South, progressively shorter belts are illuminated south and north of the light source, so that ultimately the whole pixelated ellipse is illuminated (see the bottom-right image of Fig. 2).

We now simulate this nightly elliptical illumination. The top left image of Fig. 3 depicts an isolated light source, shining at the centroid of a .56 km fine pixel located in the middle of a 21×21 grid at (row, column) = (11, 11). Without blurring, this is exactly the image we would see: only the source pixel would be illuminated; all other pixels would be dark. The top right image, by contrast, depicts the set of fine pixels (.56 km × .56 km) that are illuminated by a direct overhead pass (a so-called on-nadir pass) of a DMSP satellite. The satellite's on-nadir GSA is circular, and encompasses 18 fine pixels. The bottom left image superimposes the top right image's pixels on the larger, elliptical set of fine pixels illuminated when the satellite passes at an easterly (westerly) displacement of 750 km. In an effort to limit distortions, NOAA generates its annual composite images of nighttime lights using displacements of no more than 750 km off-nadir for each pixel. The green ellipse in the inner right image is therefore the largest ellipse (major radius = 2.54 km, minor radius = 1.88 km) used in NOAA's composites. Over the course of a year, a fixed light source may be viewed from many different displacements, 0–750 km. For simulation purposes, we assume that these displacements are drawn each night from a discrete uniform distribution of $\{0, 1, \dots, 750\}$ kilometers. A light source that is viewed on 70 cloud-free nights of a year, for example, would take 70 draws (with replacement) from this distribution.

2.2.2. Second source of blurring: geolocation error

It follows that, if there were zero geolocation error, the blurring function would simply be the integral of these differently sized ellipses, stacked neatly on top of each other, centered at the source pixel, in the manner suggested in Fig. 3's bottom left image. In reality, however, the satellite miscalculates the pixel's geolocation, so on each night not only is there a differently sized ellipse, but its centroid is shocked in a random compass direction. But by how much? In a fascinating experimental study by Tuttle et al. (2013), nightly estimates of the geolocation of isolated light sources are taken and plotted against the true, known geolocations of these sources. While the headline result of their study is that the scatter plot of the satellite's geolocation estimates exhibits a mean shift error of 2.9 km, we are interested instead in the standard deviation of the scatter plot. Although a standard deviation is not reported in their paper, we estimate from their Fig. 7 that it is approximately 1.0 km or roughly 2 fine pixels. In Fig. 3's bottom right image, we depict the on-nadir GSA illuminated around a light source at (row, column) = (11, 11). The satellite passes by on a second night, at a displacement of 440 km off nadir. But this time the satellite gets the geolocation wrong by 1/2 standard deviation, incorrectly calculating that the ellipse's centroid should be (10, 12). The ellipses from the two nights overlap considerably, but are not stacked directly on top of each other. We simulate nightly geolocation error as a random draw from a symmetric Gaussian with $\sigma = 1.0$ km. Of course, geolocation error also occurs among adjacent pixels inside an ellipse, meaning that in reality an ellipse will not have hard edges, but will 'smear' a little bit onto neighboring pixels.

2.2.3. Why the PSF widens: on-board data management

On a given night, assume a light source is viewed by the satellite's scanner. Photons enter the satellite's photomultiplier tube, the ensuing electron current is measured by an ADC, and assume that a voltage of 1000 is registered. The number 1000 is immediately taken to the 2/3 power (Elvidge et al., 1997), leaving us with 100. An ellipse is induced by the satellite's longitudinal displacement from the light source (in this case, 500 km), and a geolocation shock slightly misplaces the centroid of this ellipse (in this case to (row, column) = (10, 9)). The quantity of light is distributed evenly across the ellipse's pixels, as depicted in the

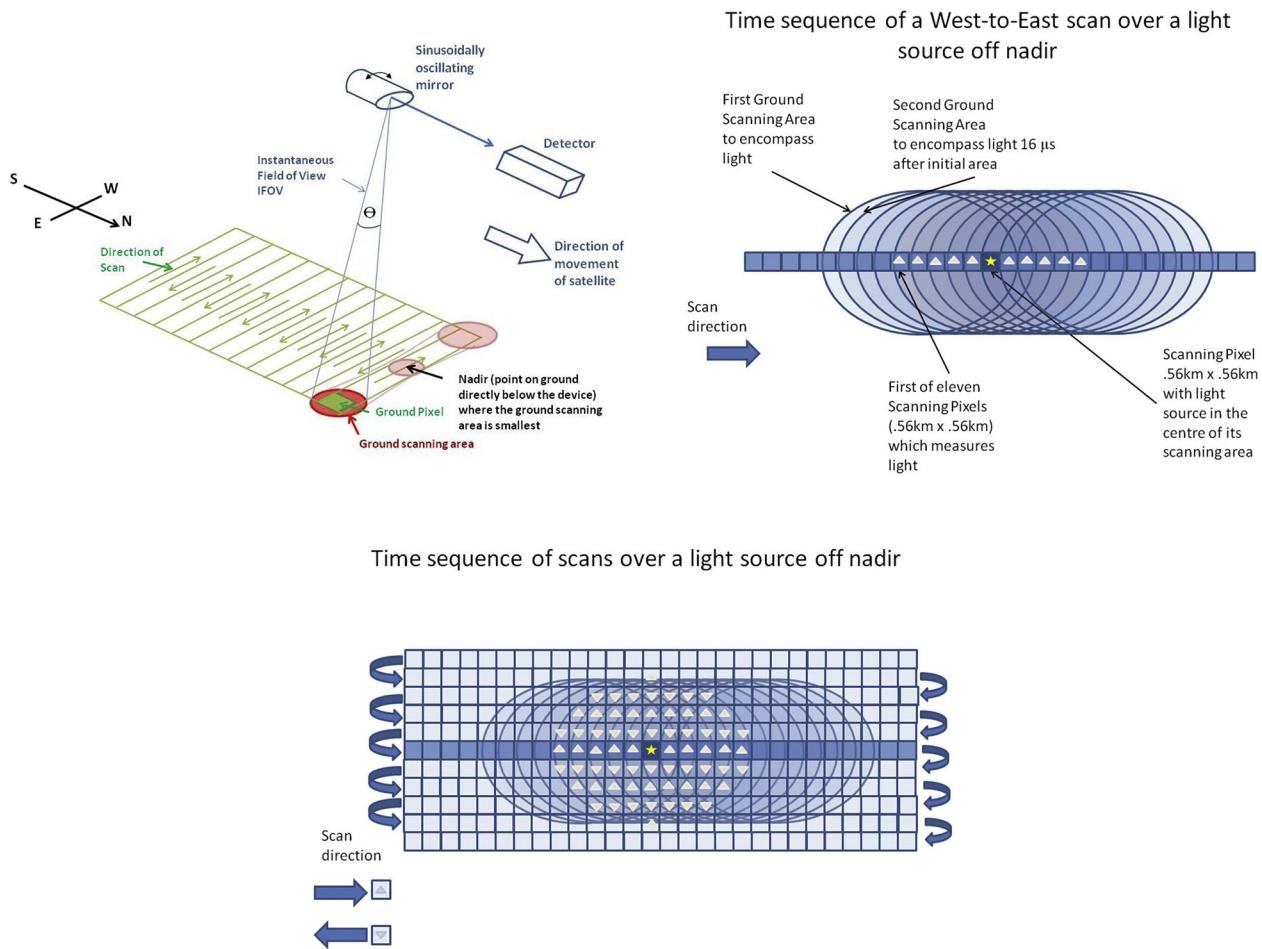


Fig. 2. DMSP-OLS data collection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

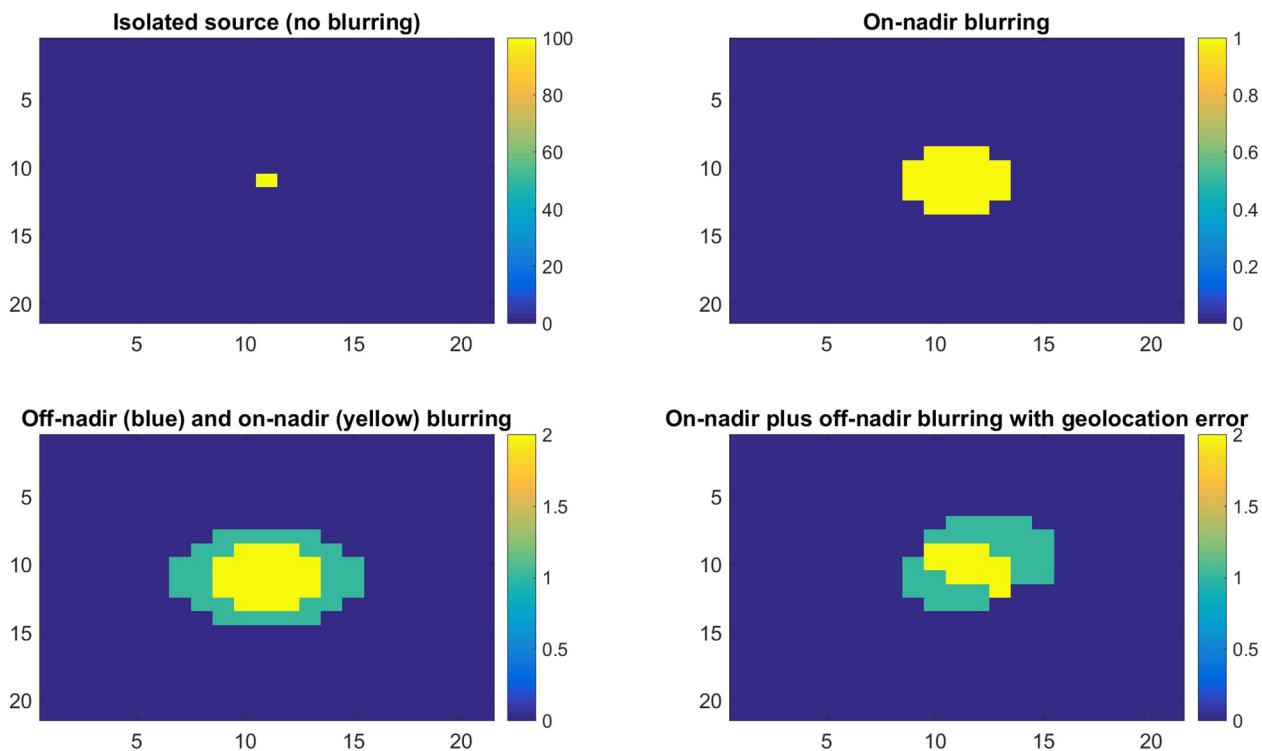


Fig. 3. Field-of-view blurring of an isolated light source. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. DMSP-NTL on-board data management.

top grid of Fig. 4. This ellipse of light smears a little bit onto neighboring pixels due to within-ellipse geolocation error (second-from-top grid in Fig. 4). Since the DMSP satellites are based on 1970s technology, however, they cannot store precise floating-point values. Instead, the numbers recorded in each fine pixel are quantized to 8 bits, meaning that all floating-point numbers are integized, and all numbers exceeding 255 are top-censored (second-from-bottom grid in Fig. 4). The DMSP satellites likewise cannot store such high-resolution information, so the resolution is coarsified: fine pixel values are summed together in consecutive 5×5 blocks, forming an image of $2.8 \text{ km} \times 2.8 \text{ km}$ coarse pixels (bottom grid in Fig. 4). Notice that the resulting coarse image will be different depending on where the first 5×5 block is drawn. A different coarse image results if the top-left corner of the first 5×5 block is at (1, 1) versus (3, 2), etc. We call the starting row the ‘row offset’, and the starting column the ‘column offset’. We treat the row and column offsets as effectively random across nights. If every night the exact same row and column offset were chosen, coarse pixels from each night would stack exactly on top of each other, and NOAA’s annual composite images would be pixelated $2.8 \text{ km} \times 2.8 \text{ km}$. Instead, however, the offsets are randomly different each night, so coarse pixels from different nights only partially overlap, recreating variation at $.56 \text{ km} \times .56 \text{ km}$ pixelization. We simulate this randomness by randomly drawing a row offset and column offset (each an integer from 1 to 5) with equiprobability every night. The bottom image in Fig. 4, for example, was generated with a (5, 1) offset, so the top-left corner of the first 5×5 block is at (5, 1), while its bottom-right corner is at (9, 5). Finally, to economize further the satellite’s on-board data storage, the smallest two bits of each coarse pixel’s value are dropped, meaning that all coarse pixel values are divided by 4, integized, and top-censored at 63, producing the 6-bit quantization familiar to users of these data.

Now that we have recreated a single-night image of this isolated light source, we can simulate NOAA’s annual composite image. Let us assume that the source was viewed on 70 cloud-free nights during the course of the year, as is typical of pixels in the Persian Gulf. We draw 70 random displacements from nadir, inducing 70 differently sized ellipses, each shocked randomly from the true centroid (11, 11) by Gaussian geolocation error, smeared by Gaussian smearing error, and coarsified at random offsets of 1 through 5. We sum these coarse images together, divide by 70 to recover the average, and integize. The resulting blur is displayed in the top left of Fig. 5.

Now that we can blur an isolated light source, we repeat this process but allow the source brightness to vary. Each time, we form East–West (x-direction) and North–South (y-direction) histograms by summing along the columns and rows of the image, respectively. We then fit normal distributions to these histograms, obtaining estimates of σ_x and σ_y in units of kilometers. We graph the results of this exercise, depicting the relationship between observed light and sizes of σ_x (bottom of Fig. 5). These graphs are the simulation equivalents of what we observed in the data in Fig. 1. By grid-search, we find $\sigma_{\text{smear}} = 0.31 \text{ km}$ and $\sigma_{\text{geo}} = 1.0 \text{ km}$ help make the simulated graphs appear most similar to the data-generated graphs.

Fig. 5’s graphs indicate that between a very low and very high threshold of light, there is a strong linear relationship between brightness of sources and PSF size. This relationship is driven by bottom- and top-censoring: although the underlying process that causes blurring is constant throughout, the quantity of light is subjected to bottom- and top-censoring through the 8-bit and 6-bit quantizations exerted on the fine and coarse pixel values. Dimmer light sources lose the tails of their PSF to bottom-censoring, where pixel values below 0.5 are recorded as 0. The PSFs of brighter light sources appear increasingly flat-topped as top-censoring decapitates their peaks at 63.

The fact that σ increases with brightness of source means that the appropriate σ with which to perform deblurring is different for every single source. The first implication of this is that we ought to choose relatively small windows to deblur, allowing σ to be larger (smaller) for brighter (dimmer) locations. But secondly, there may be locations

where there is a mixture of bright and dim sources, so the optimal σ will vary even within our window of interest. This is an unavoidable problem that ultimately places a limit on all efforts to deblur DMSP-NTL images. The optimal σ chosen for a user-defined window is in some sense the ‘average’ appropriate PSF for all light sources in the window; it will unfortunately be too small for some sources and too large for others.

2.2.4. What about atmospheric scattering?

We believe that atmospheric scattering plays a negligible role in the blurring of DMSP images. Firstly, atmospheric interference is significantly mitigated by NOAA’s in-house processing. For each pixel, NOAA exploits thermal band data to filter out cloudy nights, so the annual images are composited per pixel from cloud-free observations. Nevertheless, some residual atmospheric scattering surely remains in the images, and moreover the distribution of this scattering is generalized Gaussian, not standard Gaussian (Metari and Deschenes, 2007; we thank an anonymous referee for drawing our attention to this issue).

The contribution of this atmospheric scattering to DMSP-NTL blurring, however, is small enough to be safely ignored. To see why this is the case, consider Fig. 6. Let us assume for simplicity’s sake that on NOAA’s ‘cloud-free’ nights atmospheric particles hang suspended above the Earth’s surface at an altitude of 3 km. An Earthly light source emits light upward, but some light particles collide with the suspended atmospheric particles, and are deflected by an angle α . In reality, of course, the atmosphere has vertical dimension, so light may strike multiple atmospheric particles on its way up (*multiple scattering*). That more complicated reality, however, can be absorbed into our simpler setup by recognizing that all those deflections add to and negate from each other, summing in total to one angle α . These atmospheric collisions can therefore cause an Earthly light source to appear displaced by $d = 3 \cdot \tan(\alpha)$ kilometers. To get a sense of how large d might be on average, we need an estimate of α . Note that the same atmosphere that distorts the satellite’s view of Earthly light sources distorts our own view of celestial objects, such as the moon. The moon is just over 384,000 km from the Earth, and is roughly 3480 km in diameter. On a night with minimal cloud cover, looking up at the moon one can observe light spreading outward from the moon’s circumference to a distance of roughly 1/10th of the moon’s diameter (348 km). This implies atmospheric scattering causes a maximum angle of displacement $\alpha = \arctan(348/384,000) = 0.0009$, or 0.9 mrad. It follows that when the DMSP satellite is overhead, the light of an Earthly source may appear displaced by up to $d = 3 \cdot \tan(0.0009) = 0.0027 \text{ km}$, or just 2.7 m. DMSP-NTL pixel width is in the range of 500 m to 1 km, and observed PSFs in the DMSP-NTL images typically have a standard deviation on the order of 2 km, so a displacement of 2.7 m is orders of magnitude smaller and can safely be ignored. If we assume more generously that moonlight spreads a half moon diameter, and if we assume the cloud ceiling is 10 km instead of 3 km, we obtain $\alpha = \arctan(.5 \cdot 3840/348,000) = .0055$, in which case $d = 10 \cdot \tan(0.0055) = 0.055 \text{ km}$, or 55 m, which is still negligible relative to the nominal pixel size and typical PSF width in the DMSP-NTL images. In summary, atmospheric scattering is an insignificant source of DMSP-NTL blurring.

This analysis raises two additional questions: why does atmospheric scattering matter so much for astronomy, but so little for DMSP-NTL blurring? And if not atmospheric scattering, then what explains blurring in other nighttime satellite images such as VIIRS? To answer the first question, notice that the altitude of the atmosphere amplifies distortion. Since the atmosphere is ‘close’ to the light source, the amplification is ‘small’. Celestial light sources studied by astronomer’s, however, are far away, so amplification is large, making atmospheric scattering a major source of error for images of outer space. Regarding the second question, it is true that more modern nighttime satellites like VIIRS do a better job modulating their off-nadir GSAs and have no need to coarsify pixels for onboard data compression; but they still suffer from geolocation error. Flattening of the Earth at the poles, unevenness

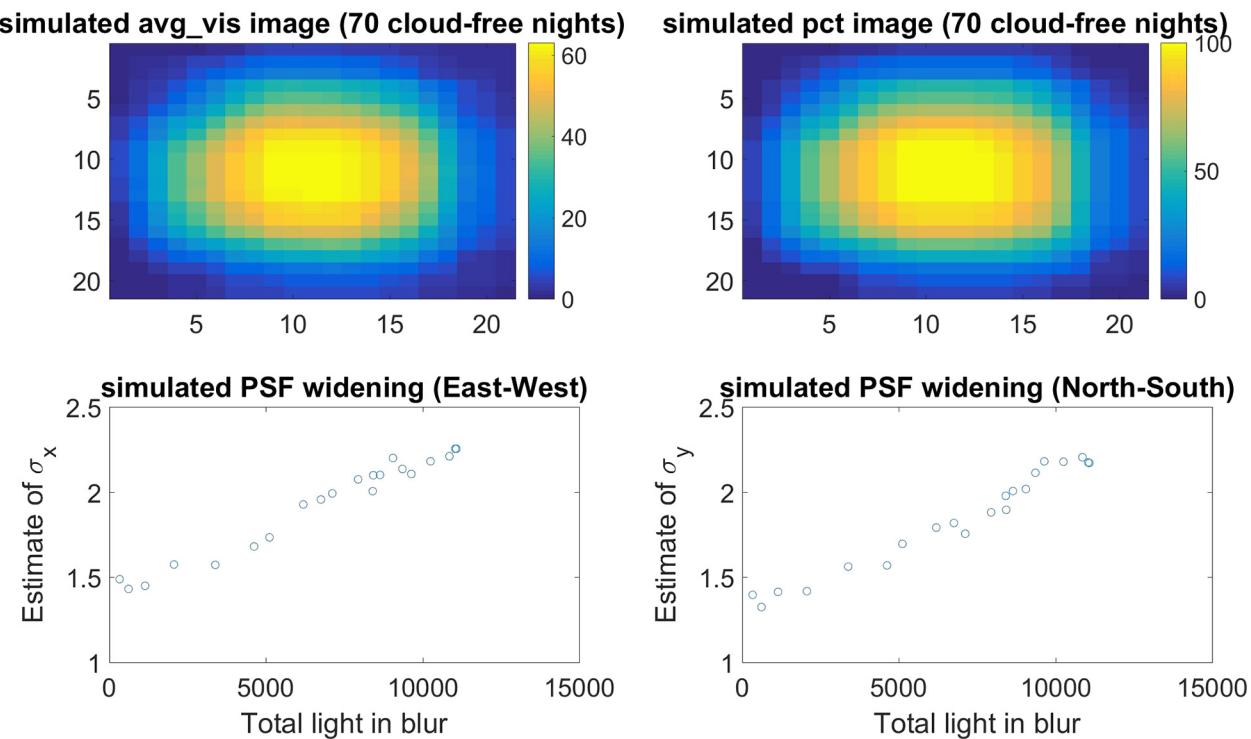


Fig. 5. Simulation results.

Simplified model with all the Atmosphere assumed to be at 3km above the earth

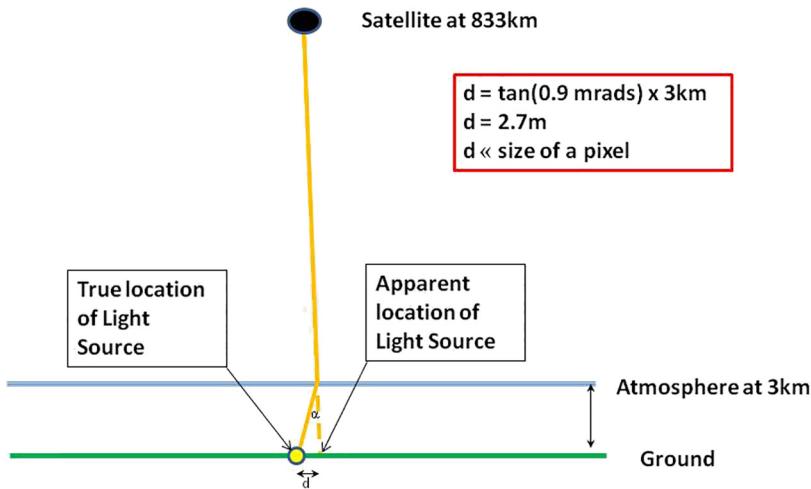


Fig. 6. Atmospheric scattering and DMSP-NTL data collection.

at the equator, and exogenous factors like solar winds all contribute to instability and precession of a satellite from its orbital track. To compensate, satellites must perform corrective thrusts that introduce geolocation error when recording images. It is beyond the scope of our study to explain blurring in VIIRS images, but we suspect that geolocation error plays a decisive role. We direct interested readers to Singer (2012), especially Fig. 3 on page 6.

2.3. Our deblurring method

2.3.1. First filter: standard deconvolution

Deblurring images with a spatially invariant PSF is a standard

problem in the image enhancement literature (see, for example, Hansen et al., 2016). The pixel values of the blurred image are stacked into one long column vector b , modeled as the true image (column vector x) left-multiplied by a blurring matrix A , yielding the equation $Ax = b$. In the absence of noise, and knowing the precise PSF, deblurring would be as simple as inverting A , obtaining $x = A^{-1}b$. In practice, however, there is always noise, and the PSF is always approximated, so even small quantities of noise can overwhelm signal when inverted. To handle noise, we employ Wiener deconvolution. One could also use a truncated spectral value decomposition (TSVD), as in Hansen et al. (2016). To use either of these methods, a noise-to-signal ratio must be specified by the user (or, equivalently, a lower bound on eigenvalues in the case of

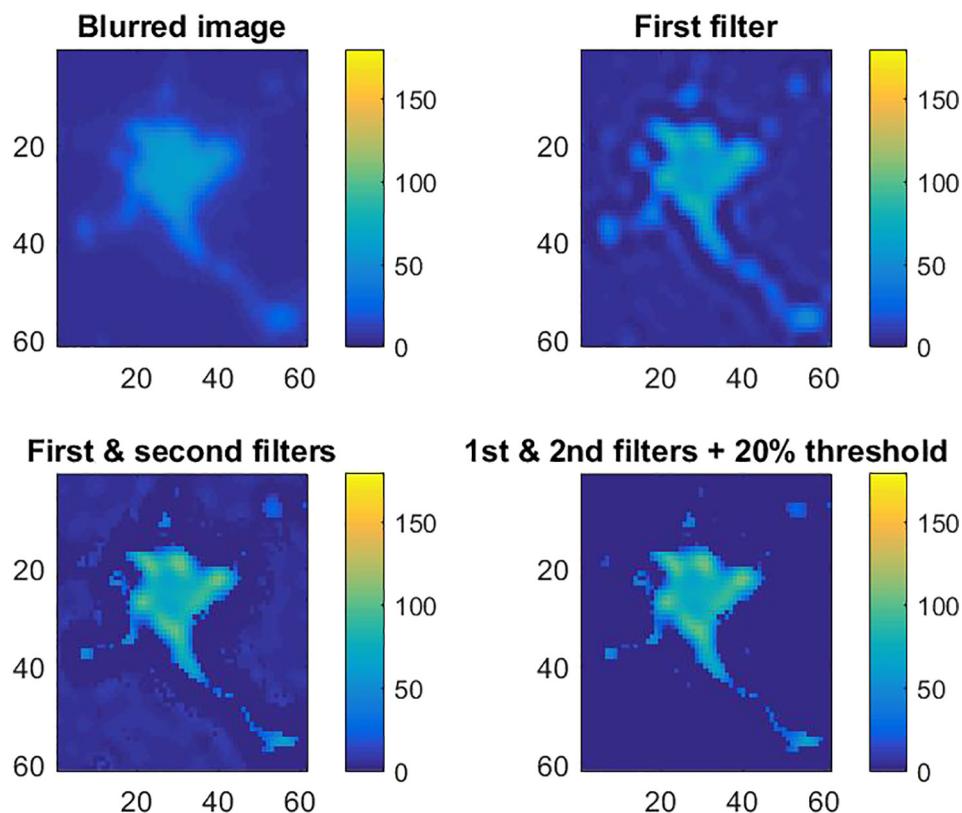


Fig. 7. Deblurring Addis Ababa, F18-2010 with $\sigma^* = 2.2$ km.

TSVD). To choose a default noise-to-signal ratio for our method, we estimate the variance of peripheral pixel values in our sample of 47 isolated sources $V ar(\text{noise})$, then estimate the variance of centrally located pixel values in these same images $V ar(\text{signal})$. Under the assumption that $Cov(\text{signal}, \text{noise}) = 0$, we recover $V ar(\text{noise})/V ar(\text{signal}) = 0.011$ for our sample.

2.3.2. Second filter: frequency of illumination maxima

For a given window, how should we choose the best σ^* for the Gaussian PSF? To answer this question, note that pixels where light sources are located (source pixels) will tend to be lit more often than neighboring non-source pixels. Consider, for example, the blurring caused by variation in the satellite's GSA. Though at each angle of displacement, the satellite records a differently sized ellipse of illumination, nevertheless the light source is always at the center of all of those ellipses. Therefore if the light source is lit $x\%$ of the time, the pixel containing the light source will likewise be recorded as lit $x\%$ of the time. A nearby pixel, by contrast, is lit only when the elliptical GSA is large enough to contain it and the source simultaneously. But ellipses will only be this large when the satellite is sufficiently far off-nadir, and this occurs only a portion of cloud-free nights of the year. As a result, even if the light source is lit $x\%$ of the time, nearby pixels will be lit $\leq x\%$ of the time. For different reasons, geolocation error predicts the same result. In this case, the entire ellipse is erroneously shifted north, south, east, or west of the source. Nearby pixels are therefore lit whenever the geolocation error shifts the ellipse in their direction, but otherwise remain dark. This again predicts that the light source is lit at least as often as its neighbors. The top right image in Fig. 5 depicts a pct image generated by our simulator, supposing that the satellite views the same point-source on 70 cloud-free nights, and that the source emits a constant 1000 units of light every night. The source pixel at (row, column) = (11, 11) is lit 100% of cloud-free nights, as are its immediate neighbors. The source pixel is therefore a local maximum in the pct image: no neighboring pixels are lit more frequently.

In summary, pixels where light sources are located are always local maxima in the pct image. The contrapositive implication is that if a pixel is not a local maximum in the pct image, then it cannot be the location of a light source, and so any light recorded at that pixel is erroneous. Our second filter, therefore, is to turn off (set to zero) all pixels in the avg_vis image that are not local maxima in the pct image. This approach differs from that of Imhoff et al. (1997) and Henderson et al. (2003), where a single, hard threshold was used (say, pixels lit less than 80% of cloud-free nights were turned off). Instead, we decide whether or not to turn off a pixel by comparing its frequency of illumination to the frequency of illumination of its eight neighboring pixels. Allowing for a 5% margin of error in percentages (arising, say, from differential cloud cover), we classify a pixel as *not* a local maximum if its percentage is exceeded by at least one of its neighbors by at least 5%. We then turn off all pixels that are not local maxima.

But this second filter ought not to be necessary. If the correct σ were applied in the first filter, then all blurred light should already have been correctly withdrawn to its true source pixels. The fact that there is light left over in non-source pixels is partly an inevitable consequence of noise, but also potentially due to an overly small or overly large choice of σ . To choose the optimal σ , we can loop back to the original blurred image and apply our first filter with a different choice of σ , then reapply the second filter. We can repeat this process, obtaining the sum of residual light among non-source pixels for each of, say, $\sigma = 1.0$ km, 1.1 km, ..., 4.0 km (recall from Fig. 1 that σ lies approximately in the range of 1.5 km to 3.0 km). The value σ^* that minimizes the sum of residual light among non-source pixels is the optimal choice for this particular choice of window, satellite, and year. Once the optimal σ^* is identified, we can apply the first and second filters in sequence and obtain the deblurred window.

3. Results

3.1. Addis Ababa

As a first demonstration of our method, let us define a window encompassing the Ethiopian capital of Addis Ababa in the image F18-2010. The (blurred) avg_vis rendering is observed in the top-left image of Fig. 7. To minimize censoring error inevitably created by the window's borders (see Hansen et al., 2016 for discussion), we actually deblur a larger window than what is depicted in Fig. 7, but we only trust the ‘interior’ solution depicted in Fig. 7. Applying our first filter with $\sigma = 2.2$ km, we obtain the top-right image. Applying the second filter to this result, we obtain the bottom-left image. This is our deblurred rendering of all light sources. It turns out, however, that most of the light sources outside the city are transitory, perhaps the result of fires or passing cars. So in the bottom-right image we apply a hard threshold, turning off all pixels whose frequency of illumination is under 20%. Notice that this action has virtually no effect on the city's appearance; the low-frequency light sources are almost all in the countryside. Depending on the particular topic of research that users are pursuing, they may wish to apply a lower or higher threshold than 20%, or perhaps not apply any threshold at all.

We record the sum of residual light and repeat this exercise for $\sigma = 1.0$ km,...,4.0 km. In Fig. 8 we can see the effect of deblurring using an overly large $\sigma = 4.0$ km. The deblurred image begins to look like Van Gogh's ‘Starry Night’ as distinct rings form around light sources. Ringing is a well known problem in the image processing literature, and is a tell-tale sign that the PSF is misspecified: $\sigma = 4.0$ km is too large for Addis Ababa F18-2010. We find that the sum of residual light is minimized at $\sigma = 2.2$ km (Fig. 7).

Reviewing Fig. 7, it is notable how the city's overall size is much smaller after deblurring. Indeed, the number of lit (nonzero) pixels declines from 3721 (blurred) to 3490 (first filter) to 2143 (second filter) to 494 (20% threshold), which comes out to an 87% reduction in total

lit area. The dramatic decline accomplished by applying the 20% threshold is impressive, but again this owes to the fact that virtually all of the pixels outside the city are initially lit, and virtually all of them turn out to be transitory light sources. While the city's outskirts go dark after the threshold is applied, the city itself looks basically the same pre- and post-threshold.

The first and second filters, however, do significantly impact the city's internal variation in brightness. In contrast to the blurred image, where the city is brightest at its center and grows darker towards the outskirts, the deblurred image reveals large pockets of darkness downtown, while the brightest parts of the city are peripheral. To explore the deblurred image more closely we superimpose the pixel values as a kmz file in Google Earth (Fig. 9). Using Google Earth's historical images, we cycle back to an image from 1986 and draw by hand a rough outline of the city's boundaries. In the depicted image from 25 years later, we can see that downtown Addis is the oldest part of the city. As confirmed with urban density data (see the discussion below), these downtown areas are also the most densely built-up parts of Addis. Given that urban areas are generally brighter than rural areas, the reader may expect that the most built-up, downtown parts of cities are brighter than the more sparsely built-up suburbs. Indeed, Zhang et al. (2013) partially rely on this idea, using vegetation intensity to negatively predict brightness in topcoded downtown areas. But readers should be careful here. The inverse relationship between vegetation and light, while broadly true when comparing rural versus urban areas, is not necessarily true when comparing downtown and suburban areas of cities. Like other East African capitals such as Nairobi and Dar es Salaam, “densely built, substandard dwellings characterize most of the old residential areas of Addis Ababa”, while “the most affluent areas of Addis Ababa display reasonably wide and tree-lined streets” (Situation Analysis of Informal Settlements in Addis Ababa, 2007). So while we may on the one hand expect that downtown areas of Addis display more light because they are denser and house more people, we may on the other hand expect that they actually generate less light due to their

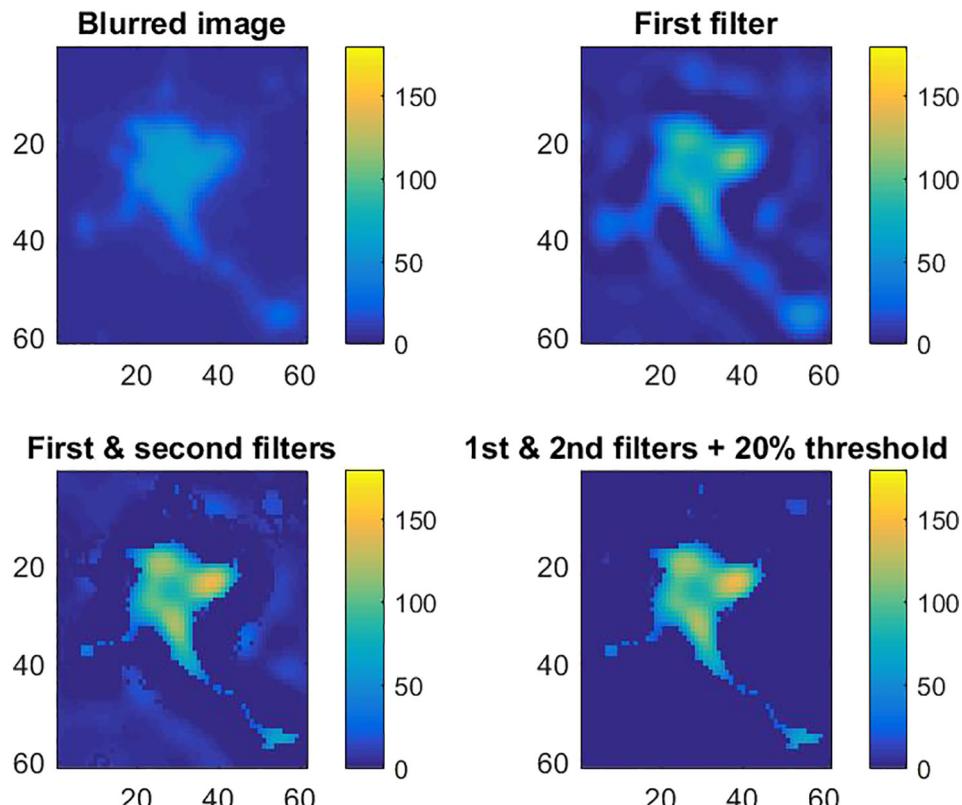


Fig. 8. Deblurring Addis Ababa, F18-2010 with overly large $\sigma = 4.0$ km.

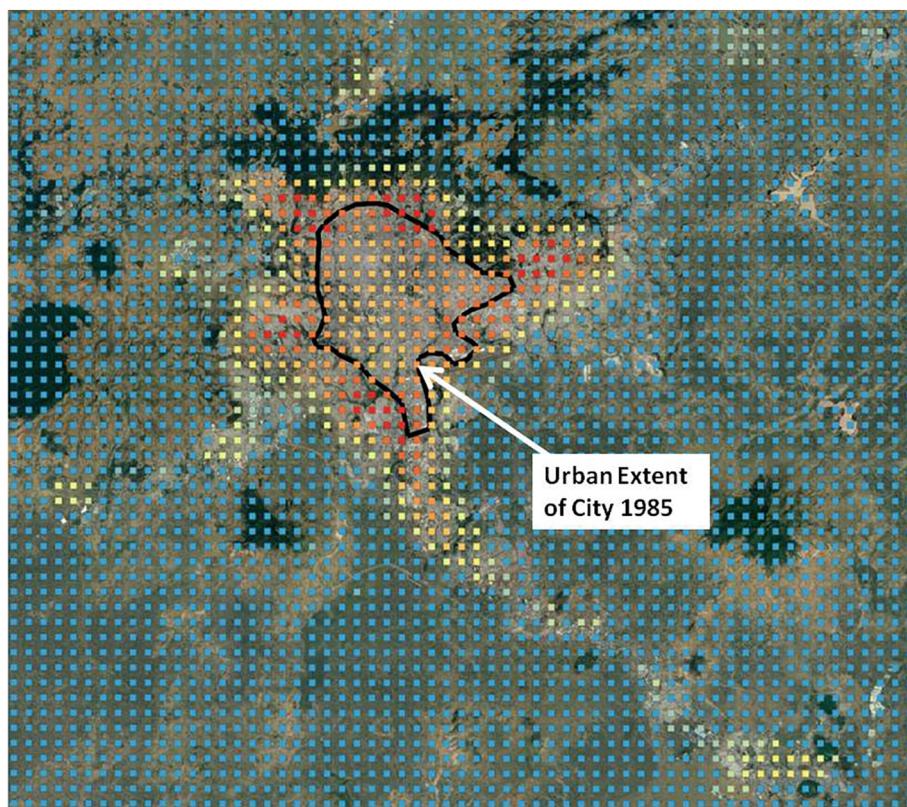


Fig. 9. Addis Ababa, F18-2010 deblurred pixels viewed in Google Earth. Blue pixels hold zero light, while yellow, orange and red pixels hold increasingly more light. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

relative impoverishment. Moreover the informality of land titling in these areas often leads to under-supply of public utilities like street-lighting: “it is well known that poor and low-income households [in Addis Ababa] do not enjoy the full benefits of electricity... the problem is worst in the squatter settlements” ([Situation Analysis of Informal Settlements in Addis Ababa, 2007](#)).

Notice finally that after the first filter's application, some pixels glow very bright, exceeding 63 counts in light. This happens because blurring caused light to spread out horizontally away from its sources, whereas deblurring reverses that process, withdrawing light back to its sources, ‘restacking’ it vertically on to its source pixels. The deblurred image exhibits greater inequalities in pixel values while the blurred image washes these away.

3.2. Hangzhou Bay Bridge

As a second demonstration of our method, we deblur the F18-2010 composite image around Hangzhou Bay Bridge, which connects the northeastern corner of China's Zhejiang province with the southern end of Shanghai province. [Fig. 10](#) offers a vertical comparison of the blurred and deblurred images. In the blurred image, every pixel is lit – the minimum pixel value is 4, the maximum is 63. In the deblurred image, there is much greater contrast. Firstly, the maximum pixel value is 118, resulting from the fact that light spilled according to the Gaussian PSF ($\sigma = 2.8$ km) has been withdrawn towards its true sources. Taking bright light sources as important areas of activity (economic or otherwise), the deblurred image helps locate these much better than the blurred image. The deblurred image also turns off many pixels that were dimly lit in the blurred image. In the bay itself, pixels are lit with low frequency, so they are turned off with our 20% threshold. The pct local maxima filter also turns off many pixels on the northern side of the frame.

Zooming in on the bridge itself, we obtain the centerline of the

bridge in Google Earth and then sum pixel values to the left and right all along the middle third of the bridge (to avoid interference from light sources at the northern or southern ends of the bridge) in both the blurred and deblurred images, obtaining the frequency plots on the underside of [Fig. 10](#). In the blurred image, the bridge is not so bright relative to adjacent pixels. Indeed, the maximum count of light actually lies just left of the bridge, since its relative dimness leaves it vulnerable to distortion by nearby sources. In the deblurred image, by contrast, the light is much more closely concentrated around the bridge's centerline, and the brightest set of pixels is that which overlaps the bridge itself.

For further comparison, we have inserted the NOAA-filtered VIIRS 2015 annual composite image (vrm-orm-ntl) of the same visual extent. Since the image is from five years later than the DMSP images, we focus our attention only on the bridge, which is effectively unchanged. We suppress the unit scale so as to prevent readers from making erroneous comparisons of brightness. VIIRS, of course, is a modern satellite with many advantages over the old DMSP series, so unsurprisingly it offers a notably sharper rendering of the bridge. Comparing the three histograms, we see that the VIIRS image concentrates more light over the bridge centerline than our deblurred DMSP image. Indeed, the standard deviation of light from the bridge centerline is just 0.52 km in the VIIRS image, compared to 0.94 km in the deblurred DMSP image and 1.99 km in the blurred DMSP image. Thus, our deblurring method improves the bridge's effective resolution by a factor of roughly two, reducing the VIIRS advantage over DMSP from a factor of four down to two.

In [Fig. 11](#) we overlay the pixel values of the blurred (top) and deblurred (bottom) image on Google Earth images (2017). Notice how the deblurred image concentrates light over major landmarks such as the Qinshan Nuclear Power Plant and the town of Haiyan. Despite the fact that the plant and town are some of the brightest areas in the image (red pixels imply > 63 counts), the *immediately adjacent* pixels in Hangzhou Bay are blue (0 counts). Our deblurring method does a near-perfect job of finding the coastline.

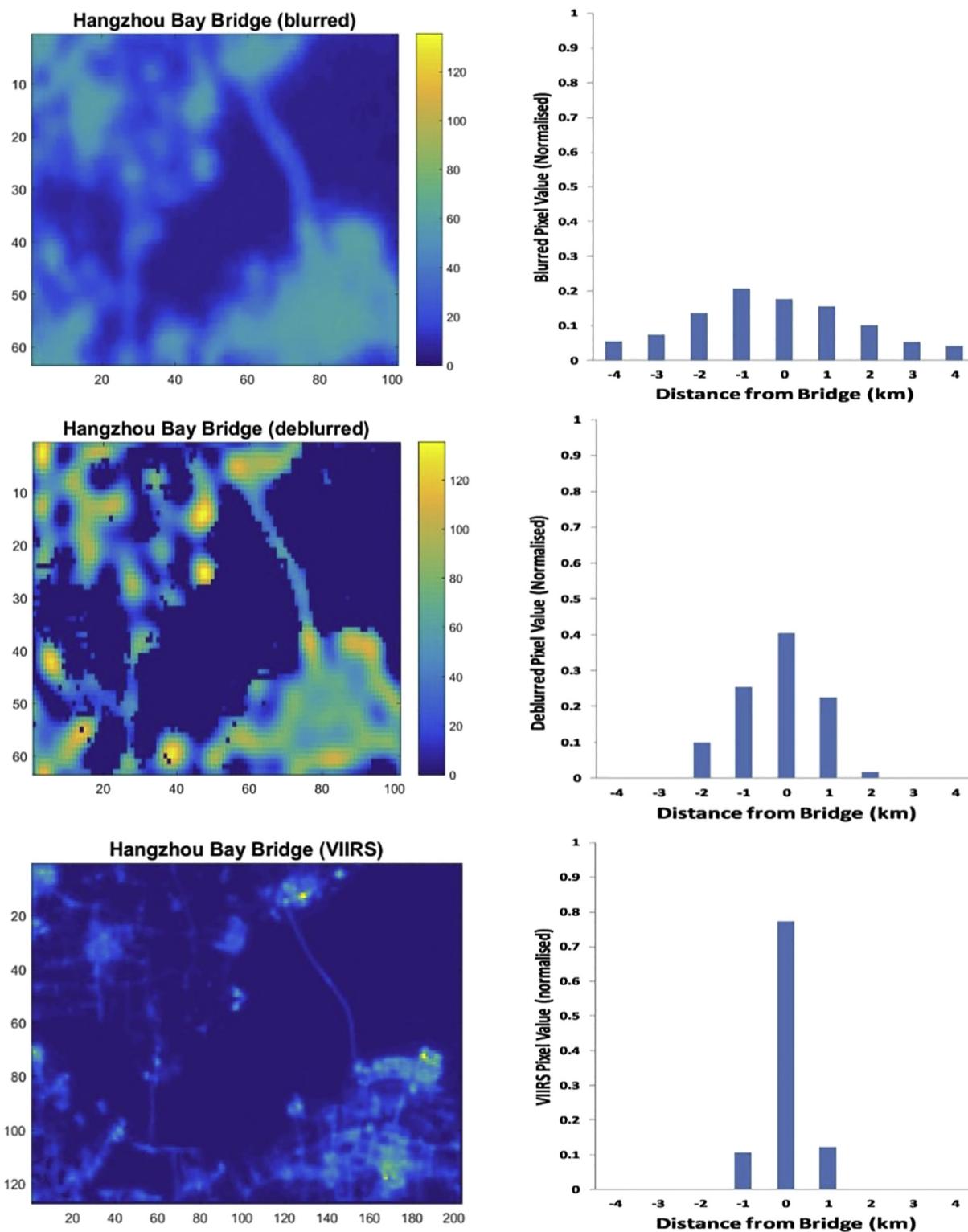


Fig. 10. Hangzhou Bay Bridge.

3.3. Isolated light sources

To provide a somewhat more generalizable evaluation of our method, we return to the F15-2000 image and our sample of 26 oil rigs in the Persian Gulf and North Sea. Fig. 12 depicts the results of deblurring the aforementioned Persian Gulf object 7 and North Sea object 4. Across the sample of 26 oil rigs, the lit area is reduced to 78.5% (after first filter), to 69.3% (after second filter), and finally to just 1.1% (after

20% threshold) of its original size.

As an alternative metric by which to quantify the improvement in resolution, we can identify the brightest pixel in each of the 26 blurred images and, on the assumption that this pixel is the oil rig's location, calculate the dispersion of light from that location before and after deblurring. For each pixel, we calculate the Euclidean distance (in kilometers) from the pixel's centroid, multiply by the pixel's quantity of light, and sum the resulting products. Performing this calculation for

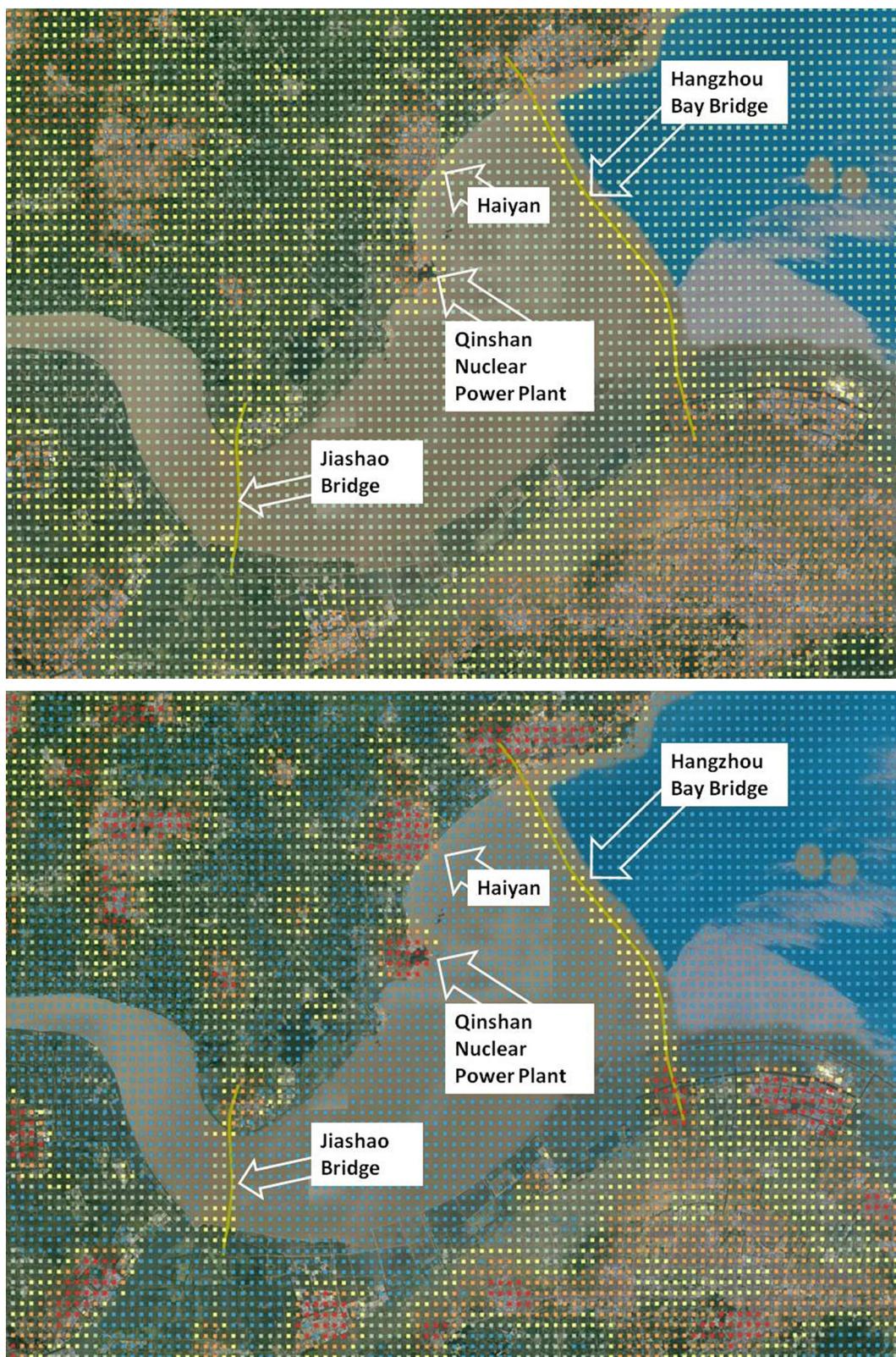


Fig. 11. Hangzhou Bay, blurred (top) and deblurred (bottom).

each of the 26 blurred images, we find that dispersion post-deblurring is just 7.5% of dispersion pre-blurring. Indeed, as visualized in Fig. 12, the deblurred image of Persian Gulf object 7 is only 10.5% as dispersed as the blurred image. The deblurred image of North Sea object 4 is 33% as dispersed as its blurred counterpart.

3.4. Urban extents

Early attempts to estimate the spatial layout of cities with DMSP images found that blurring significantly exaggerates urban extents (Imhoff et al., 1997; Henderson et al., 2003). If our deblurring

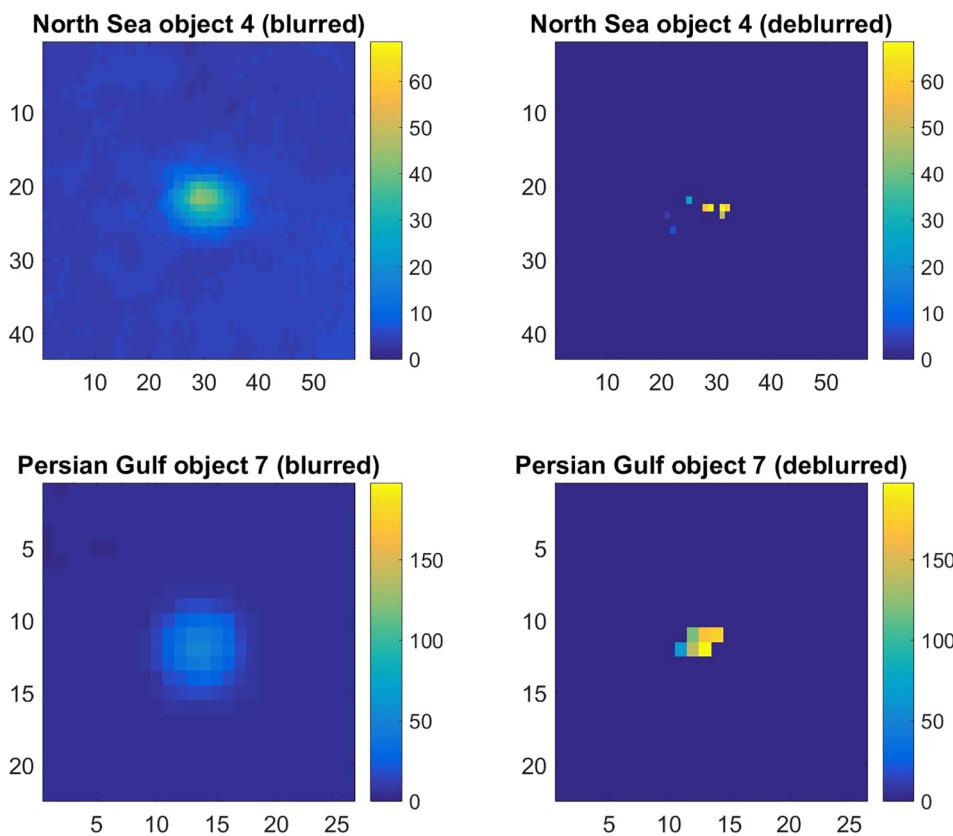


Fig. 12. Persian Gulf Object 7 and North Sea Object 4, F15-2000.

algorithm truly enhances the resolution of DMSP images, then the deblurred images ought to match urban extents much more closely than blurred images. To test this, we obtain the Global Human Settlements Layer (GHSL), a dataset created by the European Space Agency (ESA) to estimate the evolution of urban areas worldwide (Pesaresi et al., 2013). The GHSL is available for four years: 1975, 1990, 2000, and 2014. Since DMSP images are available only for 1992–2012, we use the GHSL for the year 2000. This layer relies on Landsat images from the Global Land Survey (GLS) collection organized by the Global Land Cover Facility. From these images the ESA calculates abstract radiometric and textural features. Labels for human settlements are extracted from coarse auxiliary data. Machine learning techniques then reveal the relevant abstract features predictive of urban extent. The resulting GHSL has a pixelization of approximately 39 m^2 , with each pixel labeled as either built-up or not. We access an aggregated layer whose pixelization is approximately 300 m^2 , which is still more than adequate for our purposes. Based on the fine pixels composing it, each aggregated pixel is scored by the degree to which it is ‘built up’. By default, scores are stored in 8 bits from 0 to 255, but we normalize to a scale of 0 to 100.0.

We assess the adherence of blurred versus deblurred DMSP F15-2000 images to GHSL-2000 urban extents. Since GHSL pixels are smaller than DMSP pixels, we calculate the coordinates of each GHSL pixel’s centroid, then assign GHSL pixels to their nearest DMSP pixel as determined by centroid distance. We therefore obtain for each DMSP pixel i a set of GHSL pixels $G(i)$, such that i is the DMSP pixel closest to each GHSL pixel j in the set $G(i)$. We take the set’s average built-up score $GHSL_value(i) = \frac{1}{\#G(i)} \sum_{j \in G(i)} GHSL_value_j$ and interpret this as the built-upness (urbanness) of the DMSP pixel i .

In the spirit of Imhoff et al. (1997) and Henderson et al. (2003), we assume that DMSP pixels that are not at all built-up ($GHSL_value(i) = 0$) should be completely dark ($DMSP_value(i) = 0$). So whenever a non-built-up pixel is assigned a nonzero quantity of light, we conclude it is an error, or in other words an exaggeration of urban extent. We

calculate exaggeration of urban extent for a sample of 15 cities. The sample is non-random, but was chosen with several criteria in mind. Firstly, we avoided cities with any topcoding (pixel values of 63). Our deblurring algorithm can be applied equally well to radiance-calibrated (non-topcoded) DMSP images, but to remain consistent with the rest of this paper we restricted ourselves to the regular, topcoded images, and chose to avoid highly developed cities because of their prevalent topcoding. We therefore systematically avoided North American and European cities. Secondly, we chose cities that were relatively isolated from other cities by a body of water or uninhabited countryside. This was done to minimize censoring issues arising from having too much light at the edges of our user-defined windows. Finally, we chose cities from around the world to show the algorithm can be applied anywhere. Fig. 13 depicts the sample, including 4 cities from Latin America, 5 from Africa, and 6 from Asia. For ease of visualization we normalized the blurred, stable, and deblurred DMSP images of each city to a scale of 0 to 100, where 100 represents the maximum pixel value observed in that city across all three images.

Table 1 lists these cities and their corresponding images statistics. The table’s second column records the number of pixels composing each city’s user-defined window. Due to variation in city size, and an attempt in some cases to capture the larger urban system encompassing the city, we defined larger or smaller windows, averaging 23,300 pixels per window. The table’s third column records the number of non-built-up pixels lit in the blurred DMSP F15-2000 image as a percentage of built-up pixels. On average, we find that blurred images exaggerate urban extents by 76.5%. The leftmost images in Fig. 14 depict the blurred images for 4 of our 15 cities. The inner-left images, meanwhile, depict the corresponding GHSL urban extent estimates. This estimate of 76.5% is surprisingly smaller than what was found in Imhoff et al. (1997) and Henderson et al. (2003), where DMSP images exaggerated city boundaries often by a factor of 10 (900%). Our estimate is far smaller for probably two simple reasons. Firstly, the GHSL is very sensitive,

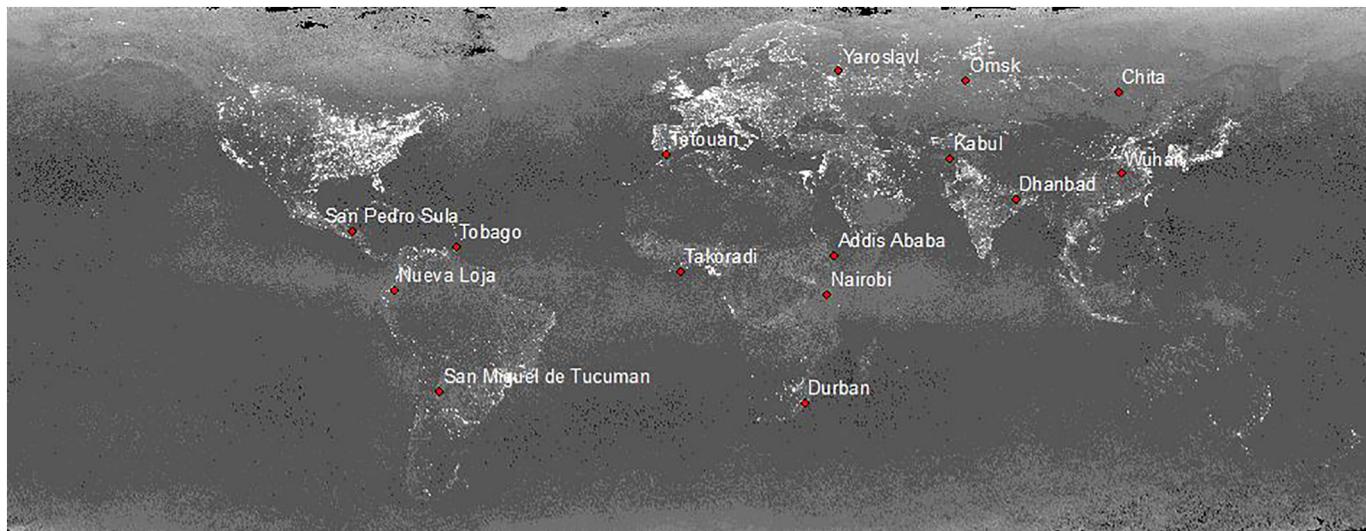


Fig. 13. Our selected sample of 15 cities for urban extent analysis.

Table I
Deblurred vs. stable vs. blurred exaggeration of GHSL urban extent.

	No. of pixels	% urban exaggeration (blurred)	% urban exaggeration (stable)	% urban exaggeration (deblurred)	Seconds to solve
Addis Ababa	4884	75.8	33.7	0.8	2.3
Chita	27,264	80.4	11.7	0.4	8.2
Dhanbad	56,950	60.3	36.4	1.1	16.0
Durban	40,414	73.1	34.9	4.7	11.7
San Pedro Sula	7140	75.2	60.6	15.6	2.8
Kabul	2773	88.1	49.1	10.8	1.6
Nairobi	2862	67.1	60.2	18.7	1.6
Nueva Loja	36,570	92.7	72.9	21.6	10.7
Omsk	22,920	77.2	66.3	10.8	7.1
San Miguel de Tucumán	20,320	76.6	49.5	7.0	6.3
Takoradi	8466	79.9	49.3	2.0	3.1
Tetouan	3015	76.4	61.5	22.2	1.6
Tobago	2852	90.0	52.4	7.4	1.6
Wuhan	93,20	59.2	29.4	0.8	26.0
Yaroslavl	19,650	75.9	65.4	8.5	6.2
Average	23,300	76.5	48.9	8.8	6.3

exploiting textural features in Landsat data that help identify even very small human settlements on the outskirts of cities. By contrast, Imhoff et al. (1997) and Henderson et al. (2003) used city administrative boundaries, which are hard thresholds that tend to ignore tiny settlements. This means we are much less likely than they to count light outside the city's formal boundaries as error. Secondly, we are extremely lenient about classifying errors. We only consider a lit DMSP pixel to be an error if its corresponding GHSL value is zero; if its GHSL value is even slightly above zero we do not count it as an error.

Aware of the blurring problem, NOAA has always provided "stable lights" images that represent their own in-house effort to filter out erroneous light. Table I's fourth column records urban exaggeration rates for the "stable lights" images, averaging 48.9%. The rightmost images of Fig. 14 depict the corresponding stable images for 4 of our cities. The inner-right windows, meanwhile, depict each city's deblurred image. Evidently the deblurred images adhere much more closely to the GHSL urban extents than do the blurred or NOAA-filtered 'stable' images. Table I confirms this visual impression in Column 5, finding that deblurred images tend to exaggerate urban extents by just 8.8% on

average. This is roughly a 5.5-fold improvement over NOAA's stable lights filtered images, and a 9-fold improvement over the blurred images.

3.5. Implementation and computational performance

Our method is implemented in a MATLAB script freely uploaded for users worldwide (<https://github.com/alexeiabrahams/nighttime-lights>). Users are required to input an "avg_vis.tif" image and its auxiliary "pct.tif" image. Users may also modify several parameters governing the script's execution. Firstly, regarding the Gaussian deconvolution, users may choose whether to use our default noise-to-signal ratio 0.011 estimated from our sample of isolated light sources, or let MATLAB estimate a noise-to-signal ratio specifically for the user's input image. Next, users may change the frequency-of-illumination threshold by which transitory light sources are filtered out. If the user's research question pertains to transitory sources, a threshold of 0% may be appropriate. Our default setting is 20%, which by inspection seems to be appropriate for eliminating non-urban features.

To give users a sense of how long tasks will require, in the final column of Table I we record the computation times for deblurring our sample of cities. Relying on just one CPU of a quad-core laptop (2.6 GHz, Intel 7th Generation), we deblurred on average 3698 pixels per second. Addis Ababa was deblurred in 2.3 s, Nairobi in 1.6. Durban and the Natal coastline of South Africa were deblurred in 11.7 s. Our script is also organized to perform parallel processing if the user so desires. To do this, the user need only change the "whole_image" parameter from 1 to 0, and change the "num_cpus" parameter from 1 to the desired number of CPUs. The script then proceeds to chop up the blurred image into overlapping sub-windows, deblur each sub-window, discard the outer regions of each sub-window to avoid edge censoring issues, then compose the final deblurred image from the deblurred inner sub-windows. The default sub-window size is 60 pixels × 60 pixels, and the default retained inner sub-window size is 10 pixels × 10 pixels. This of course raises total computation by a factor of approximately 36, but has the virtue of allowing the PSF width to be locally optimized to each sub-window (see <https://github.com/alexeiabrahams/nighttime-lights/blob/master/README.md>).

4. Conclusion

This paper offers a new method for mitigating the blurring problem in DMSP-NTL images. In agreement with Small et al. (2005) we find that blurring is governed by a symmetric Gaussian PSF, though the

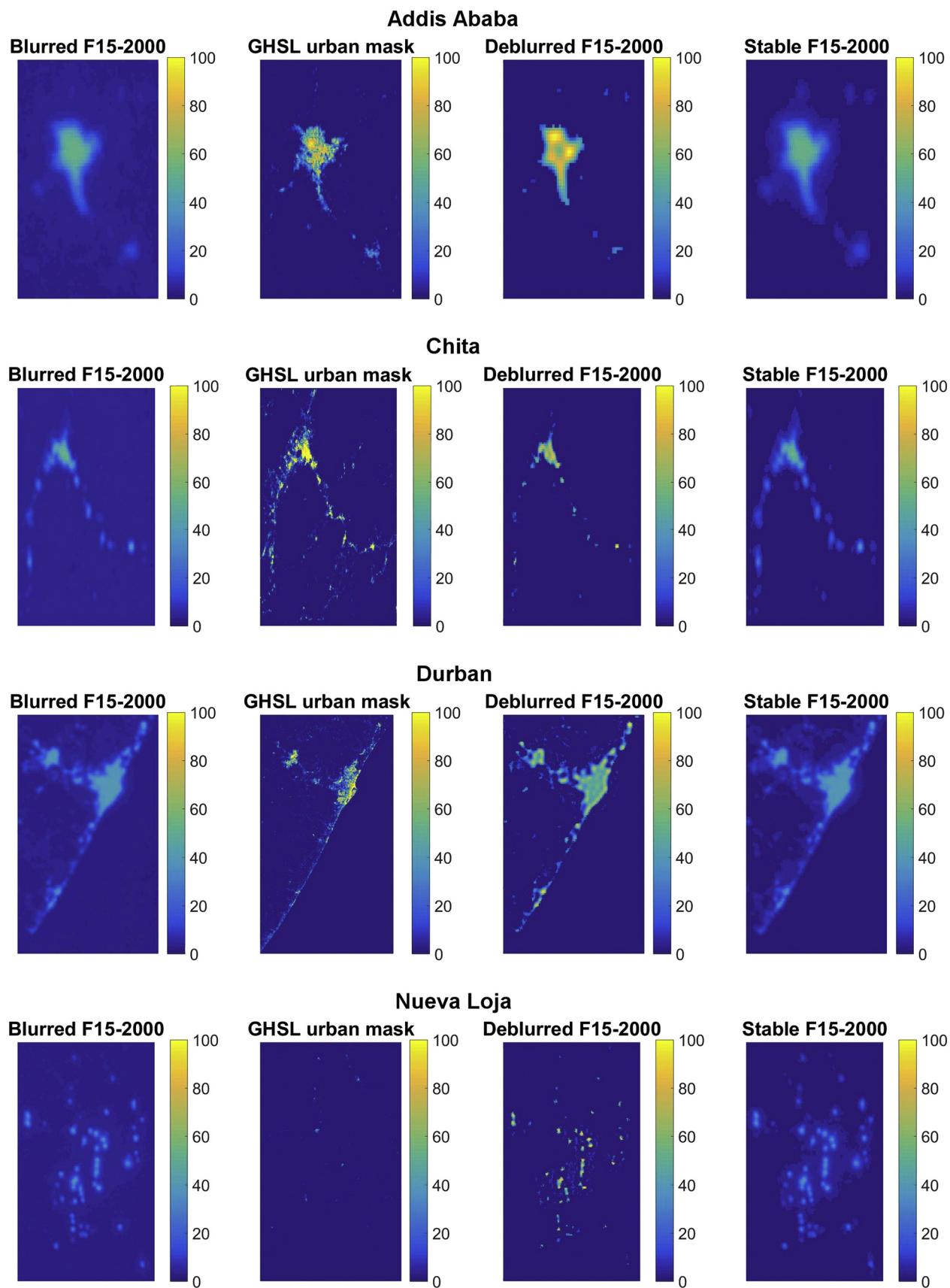


Fig. 14. Urban and DMSP comparisons for four cities.

dimensions of the PSF vary with source brightness. We discover that this relationship is driven by censoring within the satellite's onboard data management process, and we propose a way of choosing the optimal Gaussian width for a given user-defined window, satellite, and year. We implement our sequential filtering method and find that it significantly reduces the lit area around light sources. We hope that the research community will apply our MATLAB script to deblur DMSP-NTL images, and will exploit the improved resolution of the data to pursue research questions at a smaller spatial scale.

Acknowledgments

We are very grateful to David Weil and Adam Storeygard for guidance throughout the research process. We also thank C. Elvidge, J. Goldston, V. Henderson, M. Roberts, C. Small, and R. White for excellent comments. Special thanks to S. Lall for fruitful discussions and advice. This work was partially funded by UK DFID through the MDTF for Sustainable Urban Development TF071544. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development and its affiliated organizations, nor those of the Executive Directors of the World Bank nor the governments they represent. All errors are our own.

References

- Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2017. Roads, railroads, and decentralization of Chinese cities. *Rev. Econ. Stat.* 99 (3), 435–448.
- Croft, T.A., 1979. The Brightness of Lights on Earth at Night, Digitally Recorded by DMSP Satellite. US Geological Survey.
- Doll, C.N.H., 2008. Ciesin thematic guide to night-time light remote sensing and its applications. CIESIN Thematic Guide.
- Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R., 1997. Mapping city lights with nighttime data from the DMSP operational linescan system. *Photogramm. Eng. Remote. Sens.* 63, 727–734.
- Gonzalez-Navarro, M., Turner, M.A., 2016. Subways and Urban Growth: Evidence From Earth.
- Hansen, P.C., Nagy, J.G., O'Leary, D.P., 2016. Deblurring Images: Matrices, Spectra, and Filtering. Society for Industrial and Applied Mathematics.
- Henderson, M., Yeh, E.T., Gong, P., Elvidge, C.D., Baugh, K., 2003. Validation of urban boundaries derived from global night-time satellite imagery. *Int. J. Remote Sens.* 24, 595–609.
- Imhoff, M.L., Lawrence, W.T., Stutzer, D., Elvidge, C.D., 1997. A technique for using composite DMSP/OLS 'city lights' satellite data to map urban area. *Remote Sens. Environ.* 61, 361–370.
- Lieske, R.W., 1981. DMSP primary sensor data acquisition. In: Proceedings - International Telemetering Conference (U.S.). 17. pp. 1013–1020.
- Metari, S., Deschenes, F., 2007. A new convolution kernel for atmospheric point spread function applied to computer vision. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, pp. 1–8.
- Min, B., 2015. Power and the Vote: Elections and Electricity in the Developing World. Cambridge University Press.
- Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., et al., 2013. A global human settlement layer from optical HR/VHR RS data: concept and first results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 6 (5), 2102–2131.
- Pinkovskiy, M.L., 2013. Economic Discontinuities at Borders: Evidence From Satellite Data on Lights at Night.
- Sharma, R.C., Tateishi, R., Hara, K., Gharechelou, S., Iizuka, K., 2016. Global mapping of urban built-up areas of year 2014 by combining modis multispectral data with VIIRS nighttime light data. *Int. J. Digital Earth* 9 (10), 1004–1020.
- Singer, S.F., 2012. Torques and Attitude Sensing in Earth Satellites. Elsevier.
- Situation Analysis of Informal Settlements in Addis Ababa, 2007. Technical Report. In: United Nations Human Settlements Programme.
- Small, C., Pozzi, F., Elvidge, C.D., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sens. Environ.* 96, 277–291.
- Tuttle, B., Anderson, S., Sutton, P.C., Elvidge, C.D., Baugh, K., 2013. It used to be dark here. *Photogramm. Eng. Remote. Sens.* 79, 287–297.
- Zhang, Q., Schaaf, C., Seto, K.C., 2013. The vegetation adjusted NTL urban index: a new approach to reduce saturation and increase variation in nighttime luminosity. *Remote Sens. Environ.* 129, 32–41.
- Zhou, Y., Smith, S.J., Elvidge, C.D., Zhao, K., Thomson, A., Imhoff, M., 2014. A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sens. Environ.* 147, 173–185.