

Article

A Hyperspectral Survey of New York City Lighting Technology

Gregory Dobler *, Masoud Ghandehari, Steven E. Koonin and Mohit S. Sharma

NYU Center for Urban Science and Progress, 1 MetroTech Center, Brooklyn, NY 11201, USA; masoud@nyu.edu (M.G.); sek9@nyu.edu (S.E.K.); mohit.sharma@nyu.edu (M.S.S.)

* Correspondence: greg.dobler@nyu.edu

Academic Editor: Qihao Weng

Received: 26 September 2016; Accepted: 17 November 2016; Published: 5 December 2016

Abstract: Using side-facing observations of the New York City (NYC) skyline, we identify lighting technologies via spectral signatures measured with Visible and Near Infrared (VNIR) hyperspectral imaging. The instrument is a scanning, single slit spectrograph with 872 spectral channels from 0.4–1.0 μm . With a single scan, we are able to clearly match the detected spectral signatures of 13 templates of known lighting types. However, many of the observed lighting spectra do not match those that have been measured in the laboratory. We identify unknown spectra by segmenting our observations and using Template-Activated Partition (TAP) clustering with a variety of underlying unsupervised clustering methods to generate the first empirically-determined spectral catalog of roughly 40 urban lighting types. We show that, given our vantage point, we are able to determine lighting technology use for both interior and exterior lighting. Finally, we find that the total brightness of our scene shows strong peaks at the 570 nm Na-II, 595 nm Na-II and 818 nm Na-I lines that are common in high pressure sodium lamps, which dominate our observations.

Keywords: hyperspectral; lighting technology; urban science

1. Introduction

As humanity becomes progressively more urbanized, an understanding of the urban system and the interactions between its physical components—the built infrastructures, the humans who use them and the impact of that use on the natural environment—becomes progressively more important. The human/built interface is prominent in many urban subsystems, including transportation, energy and lighting use. The latter itself has significant impact on many aspects of urban functioning: various lighting technologies have different energy consumption profiles; light pollution can have significant adverse impacts on public health, including cancer risk [1,2] and disruptions of circadian rhythms [3–5]; and bright city lights can disrupt avian migration [6,7].

Observations of urban lighting have been the focus of a variety of scientific inquiries with satellite observations of nighttime urban lighting in particular having a long history. One of the earliest examples of the use of satellite imaging of urban lighting at night was [8] who used imaging from LANDSAT and DMSP (the Defense Meteorological Satellite Program) to derive correlations between nighttime images and population, urban space and energy use patterns. Later, with imagery from the Operational Linescan System (OLS; a multispectral observational platform with two broad observational bands at 0.4–1.1 μm and 10.25–12.6 μm) aboard DMSP, [9] presented methods to map spatially stable visible and near-infrared emissions from cities, while [10] explored the relationship between nighttime lighting intensity and Gross Domestic Product (GDP) for over 20 countries around the world, and [11] combined DMSP-OLS imaging with detailed U.S. census data to establish a correlation with population density at more granular scales. Since those early works, there have been a plethora of studies utilizing DMSP-OLS, as well as other imaging platforms to form ever more

detailed models of light-based proxies for economic activity [12–17], demographics and population estimates [18–21] and energy use [22–25].

Finally, two additional compelling use cases have emerged as the study of urban lighting via satellite imaging has evolved: lighting as a proxy for environmental impact and as a measure of urbanization dynamics. By exploiting the relationship between economic activity and carbon production, [12] utilized satellite detection of lighting to build a correlative model of carbon production based on satellite imaging, which [26] used to estimate carbon emissions at 1-km resolution from lighting data. To explore the potential for using nighttime satellite observations to study urbanization, [27,28] combined DMSP-OLS with LANDSAT Thematic Mapper imaging for the ground-truth to approximate the boundaries of diverse cities. Modern statistical techniques have been used to improve the robustness of the separation [29,30] (while updated LANDSAT data have provided enhanced ground-truth evaluation [31]), and there has been significant effort to use multi-modal observations to define urban boundaries [32–34]. Using long temporal baseline observations, the change in these boundaries over time has led to quantification of urban expansion in diverse areas of the globe [35–38].

While satellite observations generally give coarse views of urban lighting in both space (~ 1 km for DMSP-OLS) and time, recent work has focused on remote sensing of individual sources within an urban lightscape. The work in [39] used side-facing broadband visible observations of an urban scene to identify aggregated patterns of interior lighting use in both residential and commercial buildings. With an observational cadence of 10 seconds per image, they explored the relationship between the regular macro behavior and the more random single source behavior to quantify the “pulse” of the urban system through lighting variability.

A significant advance in the study of urban lighting came from airborne observations of Las Vegas, Nevada (USA), for which [40,41] used a ProSpecTIR-VS system with a Specim AISA Eagle VNIR and Hawk SWIR. A wavelength range of 0.4–1.4 μm at 5-nm sampling was used to detect signatures of the brightest lights used in exterior illumination. These observations were correlated with the spectra of various lighting types measured in the laboratory by the National Oceanic and Atmospheric Administration (NOAA) for the NightSat mission [42], as well as satellite maps of the terrain to pinpoint specific monuments or exterior lighting sources and assign them a type. The nature of the observations did not allow for the measurements of interior lighting types, and the spatial resolution was sufficiently coarse that only the brightest lights within a given area could be identified.

In this paper, we present the first hyperspectral survey of an urban area at high spatial and spectral sampling (~ 0.75 nm) utilizing a unique side-facing vantage point that allows us to quantify urban lighting at multiple wavelengths in a variety of settings. In Section 2, we present our observations and the methodology used to clean the data cubes from instrumental artifacts. In Section 4, we describe how we match the spectra of individual lighting technologies to known lighting types, as well as present a novel methodology for partitioning and mining the data to uncover unknown lighting types. We summarize our findings in Section 5.

2. Data Acquisition and Reduction

Figure 1 shows a daytime, composite RGB image of the field of view for our hyperspectral observations. Visible in the image are Brooklyn (near-scene), the Manhattan Bridge and East River (mid-scene), as well as the Manhattan skyline (far-scene). The same scene was imaged on the night of 20 November 2013 using a Specim Ltd. ImSpector V10E Visible Near-Infrared (VNIR) hyperspectral imager provided by Middleton Spectral Vision. The nighttime scan presented in this work was a 30-s scan, and the instrument provides 872 spectral channels from 0.4–1.03 μm with a characteristic FWHM of 0.72 nanometers. The resulting field of view is roughly $75^\circ \times 35^\circ$ with 1560×1600 pixels (the pixel axis ratio, determined by the scan rate, is ~ 0.43). The cleaned scan is shown in Figure 2.



Figure 1. A false color image of a daytime scan of the NYC skyline with our hyperspectral instrument. For this image, RGB was mapped to (610 nm, 540 nm, 475 nm). The near scene is northern Brooklyn; the mid-scene is the Manhattan Bridge and the East River; while the far scene is Midtown Manhattan.

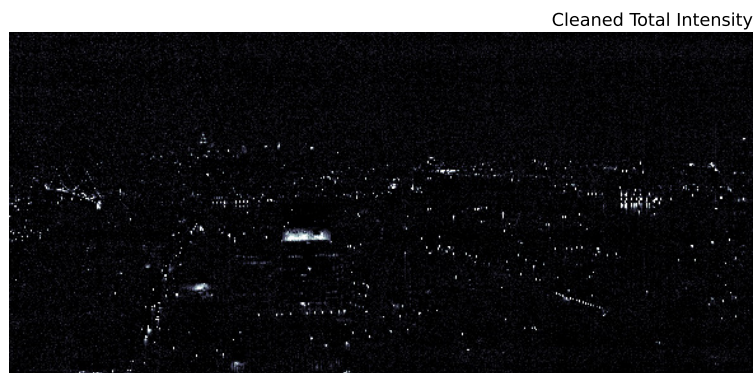


Figure 2. A night time version of the same scene as Figure 1 integrated from 0.4–1.0 μm .

2.1. Data Reduction

Figure 2 is the result of significant processing of the raw data. These raw data, integrated across all wavelengths, are shown in Figure 3. Due to the low signal to noise of the urban scene, there are strong artifacts from several effects. First, there is an overall zero-point offset between the lower and upper half of the chip. Second, gain fluctuations as the instrument scans across the scene present as vertical features in the raw image. Finally, there are smaller scale vertical artifacts below saturated pixels (a zoom in on these features is shown in Figure 4).

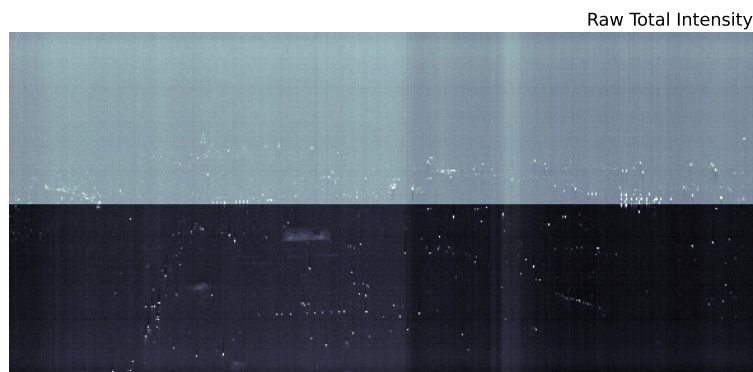


Figure 3. Raw data from the scan in Figure 2 integrated across the full 0.4–1.0- μm range. Artifacts due to chip offsets between the upper and lower half of the CCD, as well as gain changes during the scan (vertical stripes) are visible.

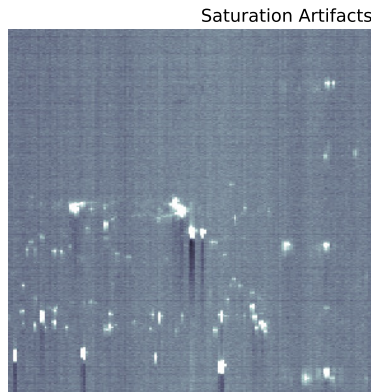


Figure 4. Saturation spikes as saturated sources are read off of the CCD.

As a first attempt to remove these artifacts from the scan, we simply subtract a dark scan from the raw data. The result is shown in Figure 5. Although the large offset between the top and bottom of the chip is removed when subtracting the dark scan, from the figure, it is clear that there are additional chip structures roughly 1/4 and 3/4 along the vertical direction that are not captured by the dark scan. These are not gain changes along the scan since those present as vertical structures in this image. Furthermore, by subtracting an unsmoothed dark model, our cleaned image will be significantly noisier than the original. From Figure 6, it is also clear that the dark current does not adequately adjust for the decreased sensitivity at short wavelengths. Therefore, we must use an alternative method to remove image artifacts.

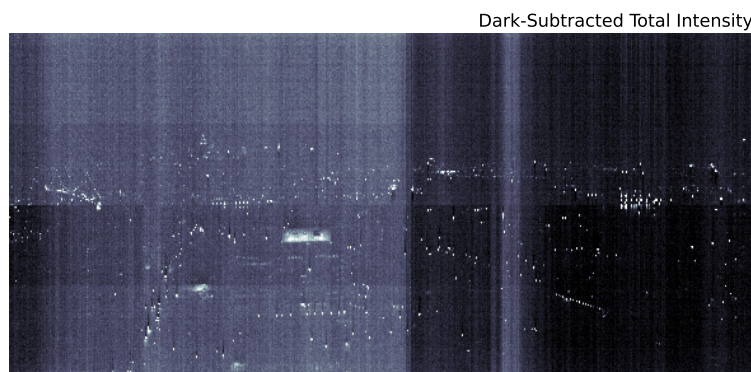


Figure 5. The same as Figure 3 except a dark scan has been subtracted. Although the chip offset is mostly removed, there are clearly additional chip features roughly 1/4 and 3/4 along the vertical direction of the CCD (the gain artifacts are obviously not removed, as well).

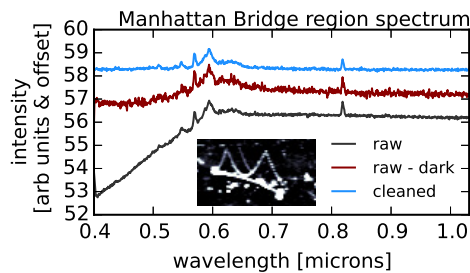


Figure 6. The raw, dark-subtracted (as in Figure 5) and cleaned (using the method described in Section 2) spectra of the Manhattan Bridge region shown in the inset. Large-scale, wavelength-dependent artifacts have been removed.

Our method is inspired by astronomical techniques for the removal of such features and handles both zero point offsets and gain changes along the scan direction. We perform the following steps for each wavelength. First, for each row i , we perform 3σ clipping by calculating the median along i , masking pixels, which are greater than 3σ (where σ is the standard deviation of pixel values along i), and calculating the median along i of the resultant values. We repeat that process 10 times and subtract the final median of i from each pixel in i . We then repeat this process across columns. The final cleaned image is shown in Figure 2. It is important to note that this method of removing artifacts relies on the image being relatively “dark”, so that a given row or column is not dominated by very bright pixels (i.e., there is a clearly-defined zero point). There is also the assumption that the background spectrum in dark regions is flat (note that ambient light pollution makes this assumption strictly incorrect; however, we find empirically that the ambient light is subdominant compared to the instrument noise).

2.2. Supplementary Data

In the following sections, we will also use external lighting spectral templates both for comparison of the results and direct correlation and regression. For the Nightsat Mission [42], the National Oceanic and Atmospheric Administration (NOAA) has made high resolution spectral measurements in the lab for various lighting types [43] (these templates were retrieved from [44]). These spectra span a wavelength range of 0.25–2.5 μm with a spectral sampling of one nanometer and are shown in Figure 7.

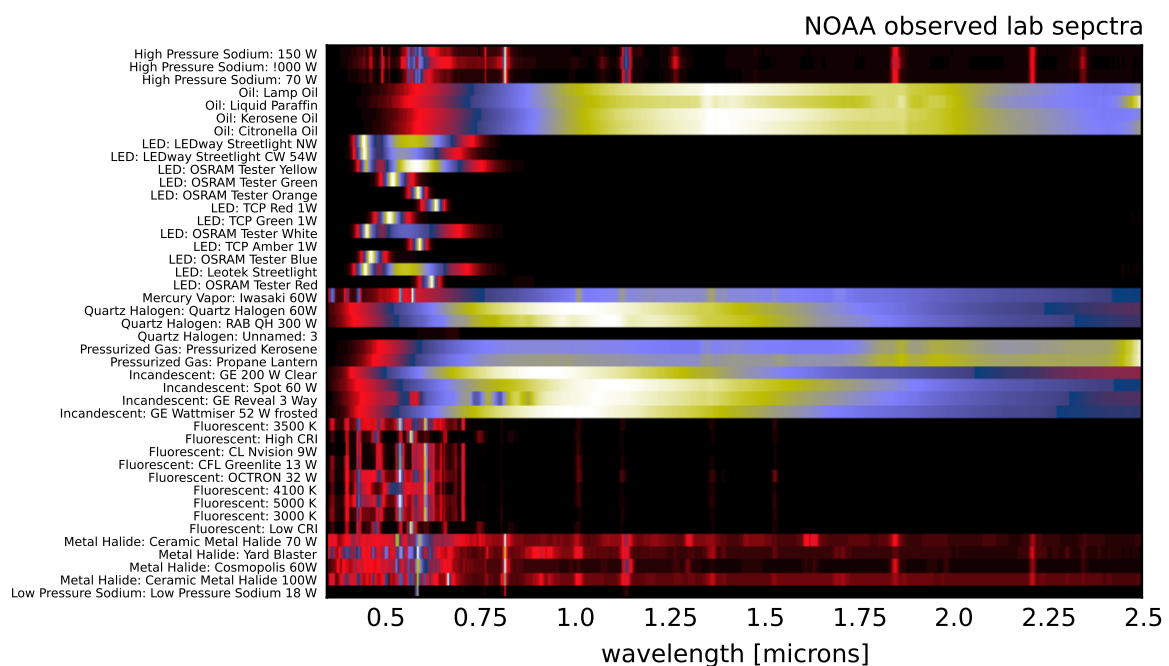


Figure 7. The 43 NOAA spectra measured by [43] in the lab. The wavelength range of 0.4–2.5 μm extends beyond the range of the instrument used in this work, though the resolution of one nanometer is comparable to our VNIR instrument.

In Section 4, we will be making comparisons between our observed spectra and the NOAA template spectra. Given that our instrument and the NOAA spectra have comparable spectral resolution, we simply interpolate the NOAA templates on to our observed wavelengths as opposed to attempting to match the frequency response of the instrument used by NOAA in the lab by convolving our observations with an appropriate bandpass filter. Not only would the impact on the analysis in the coming sections be negligible, the inherent uncertainty in any convolution kernel used to match the NOAA instrument would likely not increase our accuracy.

There is significant covariance between the NOAA spectra (particularly among multiple examples of a given lighting type). Figure 8 shows the auto-correlation matrix of the NOAA spectra after interpolating onto our observed wavelengths. The blue dots in the figure represent spectra that are correlated at greater than 90% and, thus, cannot be used to simultaneously tag spectra in our dataset. Some of these correlations are expected in that they are different spectra among the same lighting type; however, there are several strong correlations well off the diagonal (e.g., incandescent and quartz halogen spectra, which are strongly correlated with oil lantern spectra). We prune the templates by hand to represent the minimal set required for technology segregation. Our final list of 17 template spectra is shown in Figure 9.

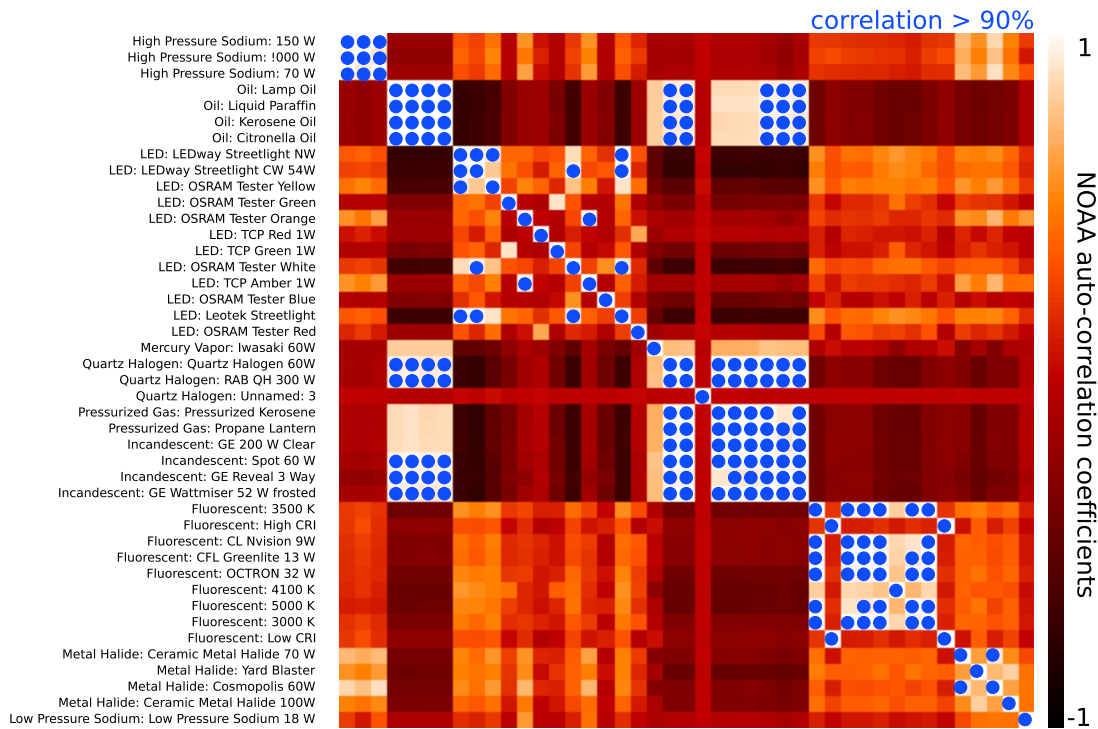


Figure 8. The auto-correlation matrix of the spectra shown in Figure 7. The high correlation between multiple examples of various lighting technologies (and in several cases between lighting technologies) implies that correlations with our spectra will only be able to identify type and not specific examples within a given type when correlating against the NOAA templates.

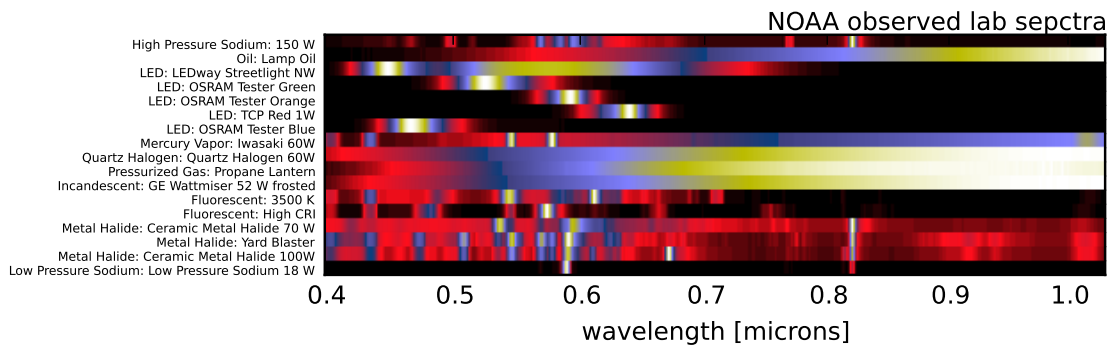


Figure 9. The result of pruning the spectra shown in Figure 7 to only include templates that have low co-variance. In addition, the templates have been interpolated onto the wavelengths of our VNIR instrument (see Section 2.2).

3. Methodology

For the vast majority of pixels in our data cube, the source brightness (either from direct or ambient light) is below our detection threshold. Therefore, to simplify the analysis and improve the stability of the algorithms we describe below, we create a signal mask in two steps. First, we create an initial mask of pixels for which the integrated (over wavelength) intensity exceeds an empirically-determined threshold value. Second, we then smooth that mask by a Gaussian of width 1 pixel and select only those pixels above a threshold of 0.25. This process strikes a balance between maintaining the edges of resolved sources, while removing isolated noise pixels (note that isolated signal pixels would be removed by this method, as well; however, we have manually checked that only ~ 50 such pixels are removed by this process). The final mask is shown in Figure 10 and clearly includes both strong signal pixels, as well as a few pixels that are dominated by noise (e.g., visible near the top of the image). Pixels that are consistent with noise will be further pruned through various methods in our analysis below. In total, there are 41,583 “active” pixels out of 2,496,000 total pixels in the image.

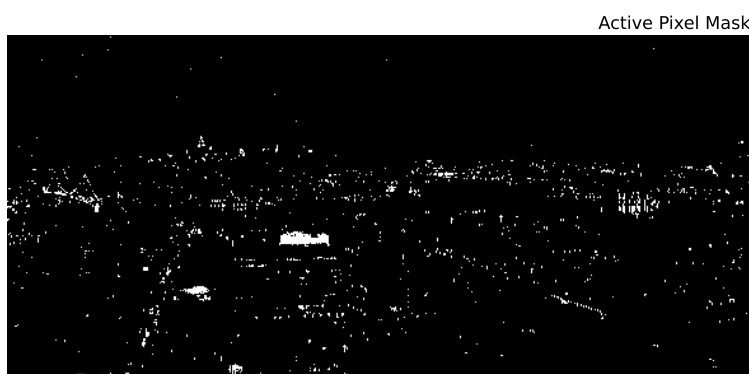


Figure 10. The pixels used in our analysis. This active pixel mask was created by thresholding the cleaned data and then thresholding a smoothed version of the result. The effect is to remove noise pixels from thresholding while maintaining the edges of extended structures.

We perform two distinct types of analysis: identifying spectral types that are known from laboratory measurements by NOAA and mining the observations for spectral types not in the NOAA sample. For the former, direct correlations, defined as:

$$C_{i,t} = \langle s'_i s'_t \rangle, \quad (1)$$

where:

$$s'_i = \frac{s_i - \langle s_i \rangle}{\sigma_i}, \quad (2)$$

s_i is the intensity as a function of wavelength for pixel i , s_t is the intensity as a function of wavelength for NOAA spectral template t , σ_i is the standard deviation of pixel i and the averages are over the wavelength, with the NOAA spectral templates suffices to identify pixels with high correspondence with known templates. For the latter, we use a variety of unsupervised machine learning and data mining techniques. Unsupervised learning has a long history of application in remote sensing in a variety of contexts (e.g., see [45–47]); however, this is the first time these methods have been applied systematically to lighting technology identification.

3.1. *k*-Means Clustering

Our first method for separating lighting types is to cluster signal pixels according to their spectral characteristics using *k*-means clustering [48–50]. The procedure first chooses k random cluster centers

(i.e., random spectra) and then, for each pixel, associates that pixel with the nearest cluster center. Here “nearest” is defined as the smallest sum of squared differences,

$$d_{i,j} = \left(\sum_m (s_i(\lambda_m) - c_j(\lambda_m))^2 \right)^{1/2}, \quad (3)$$

where $s_i(\lambda)$ and $c_j(\lambda)$ are the i -th pixel spectrum and the j -th cluster center (respectively) at wavelength λ . Each cluster center is then updated to be the mean of all $s_i(\lambda)$ associated with that cluster. The procedure is repeated until the cluster centers converge. The number of cluster centers must be set by hand prior to initialization.

3.2. Template-Activated Partition Clustering

Given the large sample size ($\sim 42,000$ spectra), it is unlikely that k -means will yield all observed lighting types for k on the order of 10. We could increase k significantly (with corresponding increases in computation time), but this would likely still miss small clusters (i.e., lighting types with only a few examples in the observations), and it is not clear that the distance-based k -means algorithm is optimal for returning “all” clusters in the data.

In fact, we note that this is an extremely difficult scenario for clustering in general for several reasons. First, the dimensionality is very high ($N_{\text{dim}} = 872$, the number of spectral channels). Second, the number of clusters is unknown. Third, the clusters themselves are not of uniform size in that different lighting types may have different spectral variability among instances. Fourth, the clusters are not uniformly sampled as there is no reason to expect that each lighting type has the same number of instances in our data. Finally, the last two imply that the cluster density is non-uniform.

Different clustering methods are useful in various scenarios (e.g., DBSCAN [51] excels at finding uniform density clusters in data), and so, we now use a combined approach of multiple clustering schemes in parallel. In particular, we use k -means, DBSCAN and hierarchical clustering [52] in tandem and filter the results to produce a reasonably orthogonal set of observed lighting types.

The analysis proceeds in three steps and is illustrated in Figure 11. First we partition the data using correlations with the NOAA templates. That is, we break the parent sample of 41,583 pixels into 10 subsamples, one that contains all spectra with $\max_t(C_{i,t}) \in (0.9, 1.0]$, one that contains all spectra with $\max_t(C_{i,t}) \in (0.8, 0.9]$, ..., and one that contains all spectra with $\max_t(C_{i,t}) < 0.1$. This Template-Activated Partition (TAP) is useful since we know that some of the NOAA templates are well represented and form significant clusters, but there are also minor variations (e.g., line ratios), which may signal different lighting types; so, those may be easiest to extract if clustered independently. In the second step, we perform the three clustering methods described above in parallel (the three clustering methods are carried out using the Scikit-learn package [53]), generating a set of cluster centers for each of the partitions and for each of the clustering methods. Finally, we filter out duplicates among all of those cluster centers by thresholding the covariance among them.

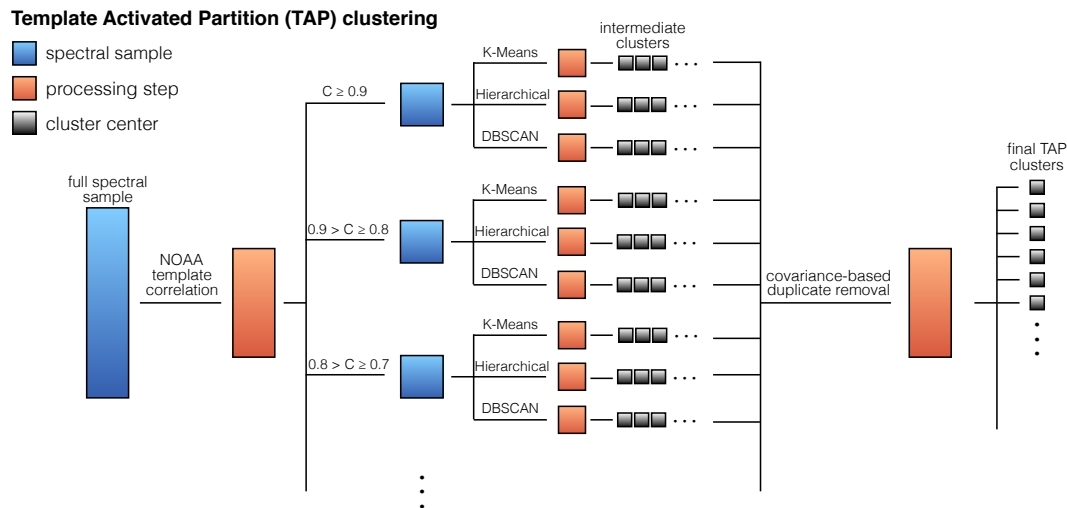


Figure 11. An illustration of the Template-Activated Partition (TAP) clustering approach. From left to right, the full spectral sample is partitioned into subsets according to the correlation with NOAA templates, each subset is independently clustered by three unsupervised clustering methods, and the results are pooled with duplicates removed to form the final TAP cluster centers.

4. Results

4.1. Correlation with Known Templates

In Figure 12, we show example pixel spectra from our scan, both for spectra that clearly correspond to the spectra measured by NOAA in the lab and also one example of a spectrum that is not in the NOAA catalog. In particular, we find high correlations (>85%) with High Pressure Sodium (HPS) lamps, various LEDs, fluorescents and metal halide lights. Figure 13 shows the correlations coefficients for all active pixels and for all NOAA templates. In our sample, we find that HPS, LED, fluorescent, metal halide and Low Pressure Sodium (LPS) lights make up the dominant lighting in our scene, while oil lamps, mercury vapor, pressurized gas, incandescent and quartz halogen lamps are largely undetected.

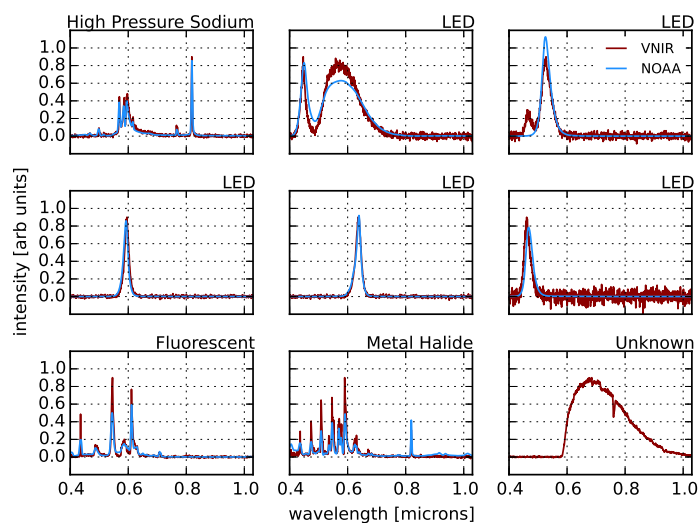


Figure 12. Several examples of high S/N spectra in our sample that are correlated with the NOAA templates shown in Figure 9. The bottom right spectrum is an example of a lighting technology that is not represented in the lab templates.

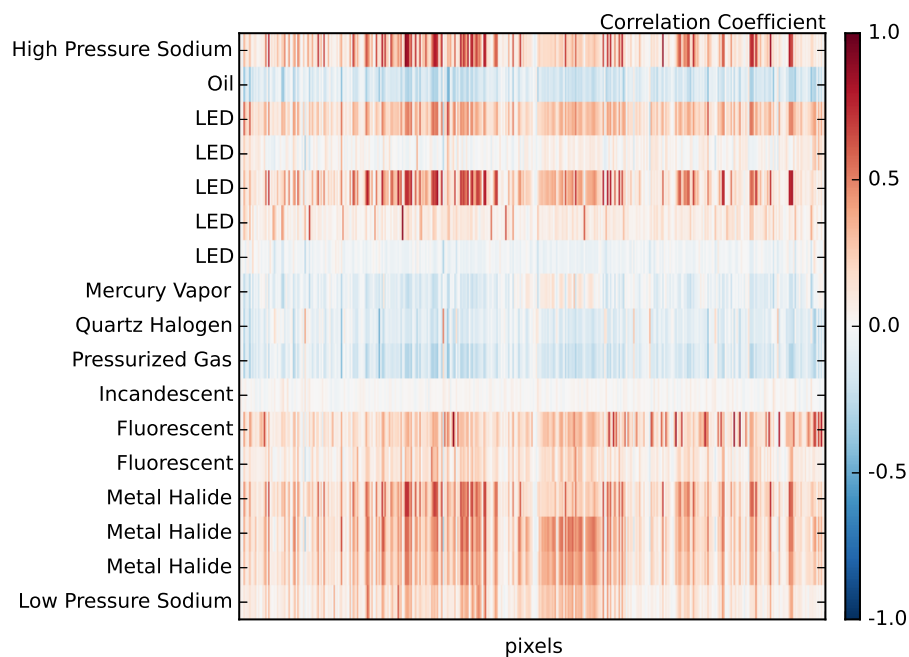


Figure 13. The correlation coefficients for all 41,583 spectra in our dataset with each of the NOAA templates in Figure 9. We clearly detect examples of 13 of the templates with high correlation coefficients. Interestingly, we do not find examples of incandescent light bulbs (see Section 4.5).

4.2. Unsupervised Learning

As we have seen from Figures 12 and 13, while several of the NOAA lighting types are detected in our scene with high confidence, many of the pixels do not correspond directly to one of those spectral templates. Nevertheless, it is reasonable to assume (and we shall see that it is indeed the case) that not all pixel spectra that are uncorrelated with the NOAA templates ($C_{i,t} < 85\%$ for all t) are unique. Rather, there are likely lighting types that are present in our sample, but not in the NOAA templates. However, there are two important unknowns, the overall number of lighting types and the number of instances of each type. We will first estimate the total number of types using unsupervised clustering techniques and then classify our pixel spectra among those.

4.2.1. k -Means Clustering

The results of applying k -means clustering to our active pixels using $k = 15$ are shown in Figure 14. Varying the number of k to be fewer or more than 15 does not result in qualitatively different cluster centers. This first pass at clustering is clearly identifying distinct lighting types that are associated with previously known spectra (see Figure 15), as well as types that are not among the NOAA templates. There are also “duplicates” among the cluster centers that represent the same type, but with minor variations in line ratios or continuum shape. We note that several of the new lighting types identified in Figure 14 are qualitatively similar to NOAA LEDs, and so, we tentatively identify their type as such, but caution that without the ground-truth, we cannot be certain.

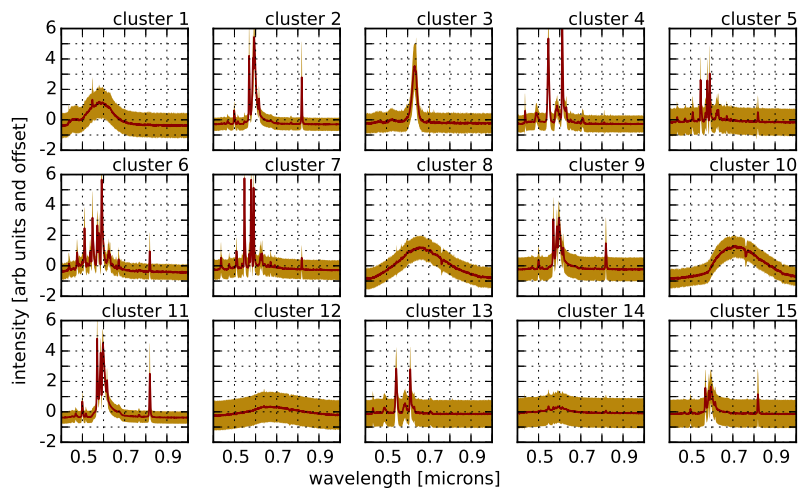


Figure 14. The result of clustering our $\sim 42,000$ spectra into 15 clusters using k -means clustering. The red lines represent the cluster centers, while the yellow bands represent the dispersion of the cluster members for each cluster. It is clear that k -means is recovering dominant spectra for High Pressure Sodium (HPS) lamps, as well as fluorescent and metal halide lamps (see Figure 15). Nevertheless, several of the cluster centers do not correspond to any NOAA templates, but are largely consistent with LED-type lights.

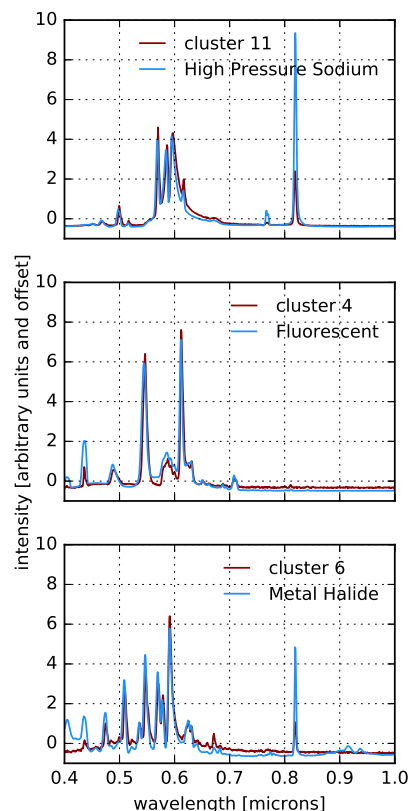


Figure 15. A comparison of several of the k -means cluster centers from Figure 14 with NOAA templates.

For lighting technologies, the ground-truth we need can come in the form of publicly available records' data. We show an example in Figure 16. The associated active pixels for six lighting types are shown in a zoom-in of the Manhattan Bridge. The HPS, fluorescent and metal halide types clearly correspond to the NOAA templates, as does one of the LED types. The remaining

two LED types are assumed to be LEDs based on their similarity to the LEDs in the top row of Figure 12. The image shows that the necklace lights of the Manhattan Bridge all correspond to the same cluster, which is presumed to be an LED. We confirm that this is indeed the case as the necklace lights of the Manhattan Bridge underwent a retrofit to LEDs shortly before our VNIR scan was taken [54]. We also point out that the lights along the bridge are HPS lamps and have been labeled by our clustering as such, while the head and tail lights of traffic along the bridge are labeled as LEDs. The figure also shows the full tagged image including the LED necklace lights on the Williamsburg bridge that were retrofitted at the same time as the Manhattan Bridge. Additional important features are the metal halide spotlights lighting the billboard near the center-left of the image, the different color LEDs in the roof versus the spire of the Empire State Building, the Chrysler building LEDs and the fluorescent interior lights in the office building directly under the billboard.

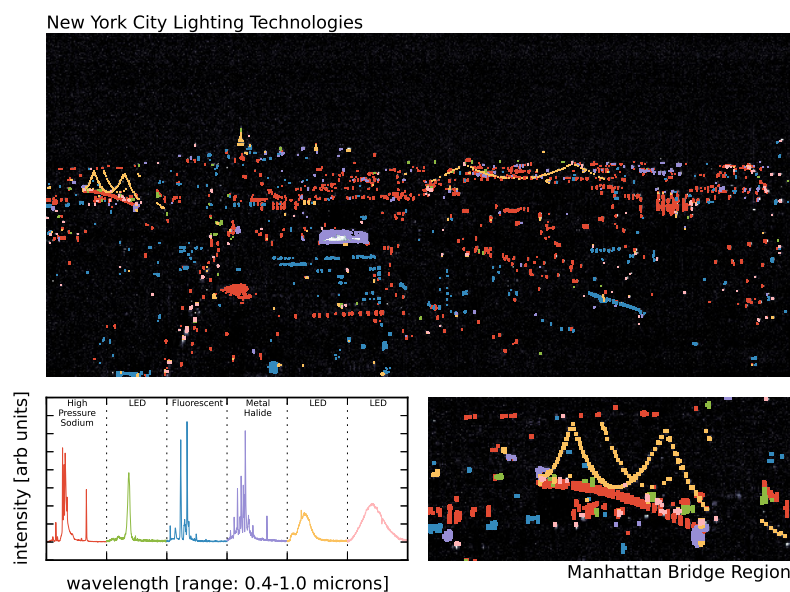


Figure 16. A projection of the cluster tags for several cluster centers on the integrated image of Figure 2 (top). The associated spectra (with corresponding colors) are shown in the lower left. A zoom-in of the Manhattan Bridge region is shown in the lower right demonstrating that detailed lighting technology use can be identified with our methodology.

4.2.2. Template-Activated Partition Clustering

While performing k -means on the full sample of $\sim 42,000$ spectra has successfully identified characteristic lighting types that are both known (from NOAA templates) and new, as noted in Section 3.2, clustering on samples of this size is not likely to generate the full set of observed lighting types with $k = 15$. Applying TAP clustering to this sample produced 42 independent clusters in the data, which are shown in Figure 17. While some of these are obviously those lighting types that are represented by the NOAA templates, there are many that are not. For example, the fourth row, second column is a cluster of 117 lights that are very bright in the far blue, but otherwise dark, while the bottom row has examples of lights with both broad and narrow line features that are beyond the human vision limit of 700 nanometers and are strongly peaked in the NIR.

These 42 different lighting types represent the first catalog of empirically-measured urban lighting technology. While incomplete, these lighting types are potentially useful in a number of urban studies, and so, we make the cluster centers, as well as the cleaned, observed spectra publicly available [55].

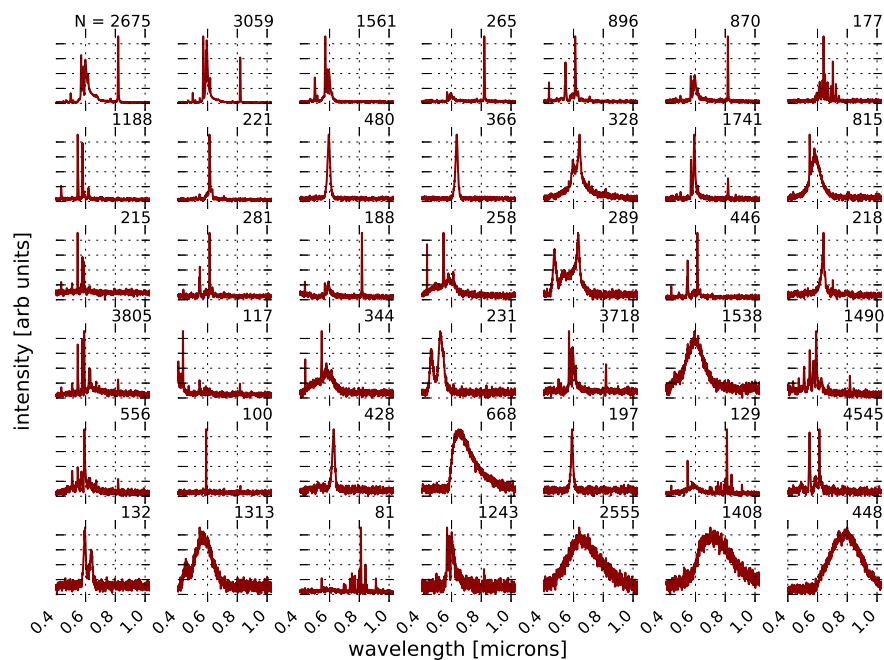


Figure 17. The results of performing template-activated partition clustering (see Figure 11 and Section 4.2.2) on our dataset to identify all statistically-robust observed lighting types. The methods of clustering used were *k*-means, DBSCAN and hierarchical.

4.3. Aggregate Spectrum

The aggregate spectrum for all pixels in our scene is shown in Figure 18. By comparing with the template spectra in Figure 12, we find that the scene is dominated by HPS lamps; and indeed, Figure 17 shows that HPS-type spectra are the most common in our dataset. In fact, the brightest lines in the aggregate spectrum are clearly Na-I and Na-II, though we also see the Hg and Eu-III lines common in fluorescent lighting.

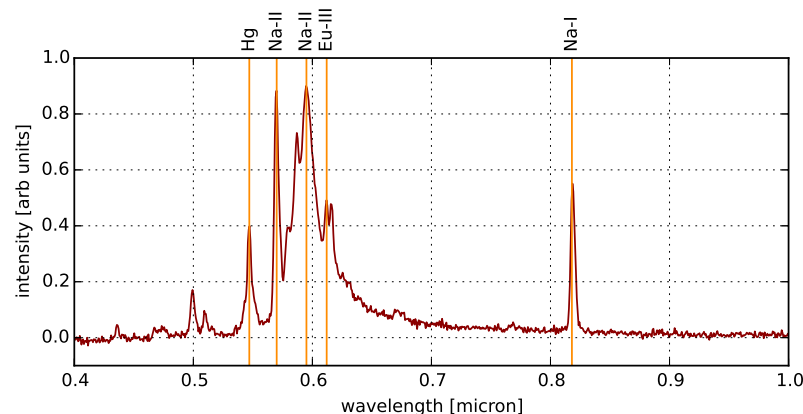


Figure 18. The aggregate spectrum of *all* pixels in our scene. The integrated spectrum of the NYC skyline is dominated by HPS-type lighting with strong peaks in the Na-I and Na-II bands, though we also identify Hg and Eu-III lines that are common in fluorescent lighting.

4.4. Applications

The sensitivity and spatial and spectral resolution of the instrument, coupled with our ability to distinguish both interior and exterior lighting technologies suggest several potential applications of our

methodology. First, with persistent scanning, we can not only measure the time-dependent brightness of the city at high spectral resolution (cf. [56] for a low spectral resolution example) for correlation with data from migratory bird strikes and human circadian rhythm disruptions (e.g., [39]), but we can assess changes in the lighting technology use over time. For example, NYC Local Law 84 requires that all large tenant spaces convert to energy-efficient lighting by 2025. In addition, given the synoptic nature of our side-facing vantage point, this technology can identify buildings for energy-efficient lighting upgrades (as in [57], but for many buildings simultaneously). Finally, since our observations of the skyline also include coverage of the sky above the city, we can assess the detailed wavelength-dependent reflectivity and amplification of artificial light by cloud cover [58,59].

4.5. Discussion

Despite the success of our methodology in differentiating various lighting technologies, there are several factors to consider before interpreting the resultant spectra. First, we note that one of the most important aspects of this work is that we have demonstrated the ability to identify lighting type for both exterior and interior lighting. For example, the rows of fluorescent lights just left of center in Figure 16 are interior lights from an office building on New York University's Tandon School of Engineering campus. While there is tremendous potential for these observational techniques as a tool to quantify technology penetration in cities, it also implies that curtains, interior reflections and mixed lighting type use (all of which are wavelength dependent) can complicate the interpretation. For the first two, sharp features in the spectra mitigate this effect to a certain extent since curtains and reflections likely have broadband impact, while for the latter, in future work, we will use component separation methods to model each lighting pixel as being mixed use.

The second complicating factor is that, although we have sufficient signal to noise in this scan to identify exterior lighting type as distant as 4–5 miles for bright building lights, for interior lights, the signal amplitude is low enough that only interior lights within about one mile are visible. Within that distance, there are no detectable incandescent light bulbs, for example. To achieve sufficient signal to noise to identify interior lighting technologies at greater distances, future work will use stacks of multiple scans.

Finally, we point out that the instrument itself does not completely cover the horizontal area across which it is scanning. As the instrument pans, it exposes the CCD at a specified frame rate, and so, horizontal areas that are scanned between exposures are not captured, potentially leading to completely missed lights. In future work, we will mitigate this effect by reducing the scan rate and exposure time and integrating pixels along the scan direction.

5. Conclusions

We have presented the first ground-based hyperspectral observations of a city at night for the purpose of empirically determining exterior and interior lighting technology usage. Our observations were carried out with a Visible Near-Infrared (VNIR) 0.4–1.0- μm single slit scanning spectrograph with 872 spectral channels. We have shown that this spectral resolution can easily distinguish lighting of various types when comparing against spectral templates generated with laboratory measurements by the National Oceanic and Atmospheric Administration (NOAA). Furthermore, the resolution and sensitivity of the instrument allows for the determination of unknown lighting types using unsupervised machine learning and data mining techniques.

Our observations are brightness limited, and by correlating pixels with external NOAA templates, we find that the dominant lighting source visible within our field of view of the New York City skyline is HPS lamps. Simple k -means clustering in $N_{\text{dim}} = 872$ dimensions of the full dataset yields cluster centers (i.e., characteristic spectra) that, in some cases, correspond to the previously known spectral templates, while in other cases representing the identification of unknown lighting types. We have used observations of the Manhattan Bridge to ground-truth and validate our unsupervised classification by

noting that our algorithm correctly identifies the bridge necklace lights as LEDs and the road lighting as HPS lamps.

We have attempted to extract as many lighting types as possible using TAP clustering. By partitioning the data into subsets based on the amplitudes of correlation with known templates and then independently clustering those subsets with a variety of unsupervised learning methods, we have identified approximately 40 different characteristic spectra (lighting types) in our scene. The clustering methods we used for the subsets (*k*-means, DBSCAN and hierarchical) were performed in parallel and then combined to yield the final catalog. By using TAP clustering to segment the data and then multiple clustering techniques that incorporate different metrics to determine cluster centers, we address the fact that clusters of lighting technologies in this space are of unknown number, of variable size, non-uniformly sampled and of variable density. Furthermore, the partitioning ensures that lighting types that have spectra with minor deviations (e.g., line ratios, continuum variations, etc.) are captured by the clustering.

One of the great promises of this methodology is to use correlative records' data to measure the relationship between urban populations and technology by empirically measuring interior lighting usage in commercial and residential scenes and correlating those with factors such as socio-economic status, demographics, business type and building age. With the present scan, we are able to detect interior fluorescent lighting for nearby (Downtown Brooklyn) office buildings, but do not have a sufficient signal to noise to detect interior lighting in the distant scene (Manhattan). However, in future work, we will stack multiple scans of the same scene to boost the signal to noise for accurate determination of interior lighting types, as well as the spectral content of diffuse light pollution in New York City.

Acknowledgments: We thank Gard Groth and Middleton Spectral Vision for their deployment of the VNIR camera and data collection. This work has been partially supported by a grant from the Alfred P. Sloan Foundation. GD's work was supported by a Complex Systems Scholar Award from the James F. McDonnell Foundation.

Author Contributions: G.D. was involved in all aspects of this work including deployment of the VNIR imaging system and data handling, analysis, and interpretation. M.G. and S.E.K. were involved in deployment of the imaging system and data interpretation. M.S.S. was involved in data handling and interpretation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schernhammer, E.S.; Laden, F.; Speizer, F.E.; Willett, W.C.; Hunter, D.J.; Kawachi, I.; Colditz, G.A. Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. *J. Natl. Cancer Inst.* **2001**, *93*, 1563–1568.
2. Kloog, I.; Stevens, R.G.; Haim, A.; Portnov, B.A. Nighttime light level co-distributes with breast cancer incidence worldwide. *Cancer Causes Control* **2010**, *21*, 2059–2068.
3. Lockley, S.W.; Brainard, G.C.; Czeisler, C.A. High sensitivity of the human circadian melatonin rhythm to resetting by short wavelength light. *J. Clin. Endocrinol. Metab.* **2003**, *88*, doi:10.1210/jc.2003-030570.
4. Navara, K.J.; Nelson, R.J. The dark side of light at night: Physiological, epidemiological, and ecological consequences. *J. Pineal Res.* **2007**, *43*, 215–224.
5. Le Tallec, T.; Perret, M.; Théry, M. Light pollution modifies the expression of daily rhythms and behavior patterns in a nocturnal primate. *PLoS ONE* **2013**, *8*, e79250.
6. Longcore, T.; Rich, C. Ecological light pollution. *Front. Ecol. Environ.* **2004**, *2*, 191–198.
7. Gauthreaux, S.A., Jr.; Belser, C.G.; Rich, C.; Longcore, T. Effects of artificial night lighting on migrating birds. In *Ecological Consequences of Artificial Night Lighting*; Rich, C., Longcore, T., Eds.; Island Press: Washington, DC, USA, 2006; pp. 67–93.
8. Welch, R. Monitoring urban population and energy utilization patterns from satellite data. *Remote Sens. Environ.* **1980**, *9*, 1–9.
9. Elvidge, C.D.; Baugh, K.E.; Kihn, E.A.; Kroehl, H.W.; Davis, E.R. Mapping city lights with nighttime data from the DMSP Operational Linescan System. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 727–734.

10. Elvidge, C.D.; Baugh, K.E.; Kihn, E.A.; Kroehl, H.W.; Davis, E.R.; Davis, C.W. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int. J. Remote Sens.* **1997**, *18*, 1373–1379.
11. Sutton, P.; Roberts, D.; Elvidge, C.; Meij, H. A comparison of nighttime satellite imagery and population density for the continental United States. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 1303–1313.
12. Doll, C.H.; Muller, J.P.; Elvidge, C.D. Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *AMBIO J. Hum. Environ.* **2000**, *29*, 157–162.
13. Doll, C.N.; Muller, J.P.; Morley, J.G. Mapping regional economic activity from night-time light satellite imagery. *Ecol. Econ.* **2006**, *57*, 75–92.
14. Ghosh, T.; Powell, R.L.; Elvidge, C.D.; Baugh, K.E.; Sutton, P.C.; Anderson, S. Shedding light on the global distribution of economic activity. *Open Geogr. J.* **2010**, *3*, 147–160.
15. Henderson, J.V.; Storeygard, A.; Weil, D.N. Measuring economic growth from outer space. *Am. Econ. Rev.* **2012**, *102*, 994–1028.
16. Li, X.; Xu, H.; Chen, X.; Li, C. Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China. *Remote Sens.* **2013**, *5*, 3057–3081.
17. Keola, S.; Andersson, M.; Hall, O. Monitoring economic development from space: Using nighttime light and land cover data to measure economic growth. *World Dev.* **2015**, *66*, 322–334.
18. Lo, C. Modeling the population of China using DMSP operational linescan system nighttime data. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 1037–1047.
19. Briggs, D.J.; Gulliver, J.; Fecht, D.; Vienneau, D.M. Dasymeric modelling of small-area population distribution using land cover and light emissions data. *Remote Sens. Environ.* **2007**, *108*, 451–466.
20. Levin, N.; Duke, Y. High spatial resolution night-time light images for demographic and socio-economic studies. *Remote Sens. Environ.* **2012**, *119*, 1–10.
21. Yang, X.; Yue, W.; Gao, D. Spatial improvement of human population distribution based on multi-sensor remote-sensing data: An input for exposure assessment. *Int. J. Remote Sens.* **2013**, *34*, 5569–5583.
22. Lo, C. Urban indicators of China from radiance-calibrated digital DMSP-OLS nighttime images. *Ann. Assoc. Am. Geogr.* **2002**, *92*, 225–240.
23. Letu, H.; Hara, M.; Yagi, H.; Naoki, K.; Tana, G.; Nishio, F.; Shuhei, O. Estimating energy consumption from night-time DMPS/OLS imagery after correcting for saturation effects. *Int. J. Remote Sens.* **2010**, *31*, 4443–4458.
24. He, C.; Ma, Q.; Liu, Z.; Zhang, Q. Modeling the spatiotemporal dynamics of electric power consumption in Mainland China using saturation-corrected DMSP/OLS nighttime stable light data. *Int. J. Digit. Earth* **2014**, *7*, 993–1014.
25. Xie, Y.; Weng, Q. Detecting urban-scale dynamics of electricity consumption at Chinese cities using time-series DMSP-OLS (Defense Meteorological Satellite Program-Operational Linescan System) nighttime light imageries. *Energy* **2016**, *100*, 177–189.
26. Ghosh, T.; Elvidge, C.D.; Sutton, P.C.; Baugh, K.E.; Ziskin, D.; Tuttle, B.T. Creating a global grid of distributed fossil fuel CO₂ emissions from nighttime satellite imagery. *Energies* **2010**, *3*, 1895–1913.
27. Henderson, M.; Yeh, E.T.; Gong, P.; Elvidge, C.; Baugh, K. Validation of urban boundaries derived from global night-time satellite imagery. *Int. J. Remote Sens.* **2003**, *24*, 595–609.
28. Small, C.; Pozzi, F.; Elvidge, C.D. Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sens. Environ.* **2005**, *96*, 277–291.
29. Small, C.; Elvidge, C.D.; Balk, D.; Montgomery, M. Spatial scaling of stable night lights. *Remote Sens. Environ.* **2011**, *115*, 269–280.
30. Zhou, Y.; Smith, S.J.; Elvidge, C.D.; Zhao, K.; Thomson, A.; Imhoff, M. A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sens. Environ.* **2014**, *147*, 173–185.
31. Shi, K.; Huang, C.; Yu, B.; Yin, B.; Huang, Y.; Wu, J. Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas. *Remote Sens. Lett.* **2014**, *5*, 358–366.
32. Lu, D.; Tian, H.; Zhou, G.; Ge, H. Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. *Remote Sens. Environ.* **2008**, *112*, 3668–3679.
33. Cao, X.; Chen, J.; Imura, H.; Higashi, O. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. *Remote Sens. Environ.* **2009**, *113*, 2205–2209.

34. Pandey, B.; Joshi, P.; Seto, K.C. Monitoring urbanization dynamics in India using DMSP/OLS night time lights and SPOT-VGT data. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 49–61.
35. Zhang, Q.; Seto, K.C. Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sens. Environ.* **2011**, *115*, 2320–2329.
36. Liu, Z.; He, C.; Zhang, Q.; Huang, Q.; Yang, Y. Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landsc. Urban Plan.* **2012**, *106*, 62–72.
37. Small, C.; Elvidge, C.D. Night on Earth: Mapping decadal changes of anthropogenic night light in Asia. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *22*, 40–52.
38. Xie, Y.; Weng, Q. Updating urban extents with nighttime light imagery by using an object-based thresholding method. *Remote Sens. Environ.* **2016**, *187*, 1–13.
39. Dobler, G.; Ghandehari, M.; Koonin, S.E.; Nazari, R.; Patrinos, A.; Sharma, M.S.; Tafvizi, A.; Vo, H.T.; Wurtele, J.S. Dynamics of the urban lightscape. *Inf. Syst.* **2015**, *54*, 115–126.
40. Kruse, F.A.; Elvidge, C.D. Identifying and mapping night lights using imaging spectrometry. In Proceedings of the 2011 IEEE Aerospace Conference, Big Sky, MT, USA, 5–12 March 2011; IEEE: New York, NY, USA, 2011; pp. 1–6.
41. Kruse, F.A.; Elvidge, C.D. Characterizing urban light sources using imaging spectrometry. In Proceedings of the 2011 Joint Urban Remote Sensing Event, Munich, Germany, 10–13 April 2011; IEEE: New York, NY, USA, 2011; pp. 149–152.
42. Elvidge, C.D.; Cinzano, P.; Pettit, D.; Arvesen, J.; Sutton, P.; Small, C.; Nemani, R.; Longcore, T.; Rich, C.; Safran, J.; et al. The Nightsat mission concept. *Int. J. Remote Sens.* **2007**, *28*, 2645–2670.
43. Elvidge, C.D.; Keith, D.M.; Tuttle, B.T.; Baugh, K.E. Spectral identification of lighting type and character. *Sensors* **2010**, *10*, 3961–3988.
44. Earth Observation Group—Defense Meteorological Satellite Program, Boulder | [ngdc.noaa.gov](https://www.ngdc.noaa.gov/eog/night_sat/nightsat.html). Available online: https://www.ngdc.noaa.gov/eog/night_sat/nightsat.html (accessed on 30 November 2016).
45. Richards, J.A.; Richards, J. *Remote Sensing Digital Image Analysis*; Springer: Heidelberg, Germany, 1999; Volume 3.
46. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*; Academic Press: Cambridge, MA, USA, 2006.
47. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36.
48. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 27 December 1965–7 January 1966; Volume 1, pp. 281–297.
49. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.
50. Celebi, M.E.; Kingravi, H.A.; Vela, P.A. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Syst. Appl.* **2013**, *40*, 200–210.
51. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.
52. Ward, J.H., Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.
53. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
54. NYC DOT—Press Releases—Expansion of Energy-Efficient LED-Light Installations Citywide. Available online: http://www.nyc.gov/html/dot/html/pr2012/pr12_19.shtml (accessed on 30 November 2016).
55. Physical Sciences Center for Urban Science. Available online: <http://serv.cusp.nyu.edu/~gdobler> (accessed on 30 November 2016).
56. Puschnig, J.; Posch, T.; Uttenthaler, S. Night sky photometry and spectroscopy performed at the Vienna University Observatory. *J. Quant. Spectrosc. Radiat. Transf.* **2014**, *139*, 64–75.
57. Venable, K.; McGee, K.; McGarry, D.; Wightman, K.; Gutierrez, C.; Weiser, R.; Vanderford, B.; Knight, J.; Rasmussen, B.; Coverick, R.; et al. Image Recognition System for Automated Lighting Retrofit Assessment, 2013. Available online: <http://oaktrust.library.tamu.edu/handle/1969.1/149181> (accessed on 30 November 2016).

58. Kyba, C.C.; Ruhtz, T.; Fischer, J.; Hölker, F. Cloud coverage acts as an amplifier for ecological light pollution in urban ecosystems. *PLoS ONE* **2011**, *6*, e17307.
59. Kyba, C.; Ruhtz, T.; Fischer, J.; Hölker, F. Red is the new black: how the colour of urban skyglow varies with cloud cover. *Mon. Not. R. Astron. Soc.* **2012**, *425*, 701–708.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).