

## Dynamics of the urban lightscape



Gregory Dobler<sup>a,\*</sup>, Masoud Ghandehari<sup>a</sup>, Steven E. Koonin<sup>a</sup>, Rouzbeh Nazari<sup>a,b</sup>,  
Aristides Patrinos<sup>a</sup>, Mohit S. Sharma<sup>a</sup>, Arya Tafvizi<sup>a</sup>, Huy T. Vo<sup>a</sup>,  
Jonathan S. Wurtele<sup>c</sup>

<sup>a</sup> Center for Urban Science and Progress, New York University, New York, NY 11201, United States

<sup>b</sup> Department of Civil & Environmental Engineering, Rowan University, Glasboro, NJ 08028, United States

<sup>c</sup> Department of Physics, University of California at Berkeley and Lawrence Berkeley Laboratory, Berkeley, CA 94720, United States

### ARTICLE INFO

#### Article history:

Received 30 November 2014

Received in revised form

13 May 2015

Accepted 3 June 2015

Available online 30 June 2015

#### Keywords:

Urban imaging

Visible light observations

Time series

Image processing

Pattern recognition

### ABSTRACT

The manifest importance of cities and the advent of novel data about them are stimulating interest in both basic and applied “urban science” (Bettencourt et al., 2007 [4]; Bettencourt, 2013 [3]). A central task in this emerging field is to document and understand the “pulse of the city” in its diverse manifestations (e.g., in mobility, energy use, communications, economics) both to define the normal state against which anomalies can be judged and to understand how macroscopic city observables emerge from the aggregate behavior of many individuals (Louail, 2013 [9]; Ferreira et al., 2013 [6]). Here we quantify the dynamics of an urban lightscape through the novel modality of persistent synoptic observations from an urban vantage point. Established astronomical techniques are applied to visible light images captured at 0.1 Hz to extract and analyze the light curves of 4147 sources in an urban scene over a period of 3 weeks. We find that both residential and commercial sources in our scene exhibit recurring aggregate patterns, while the individual sources decorrelate by an average of one hour after only one night. These highly granular, stand-off observations of aggregate human behavior – which do not require surveys, *in situ* monitors, or other intrusive methodologies – have a direct relationship to average and dynamic energy usage, lighting technology, and the impacts of light pollution. They may also be used indirectly to address questions in urban operations as well as behavioral and health science. Our methodology can be extended to other remote sensing modalities and, when combined with correlative data, can yield new insights into cities and their inhabitants.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

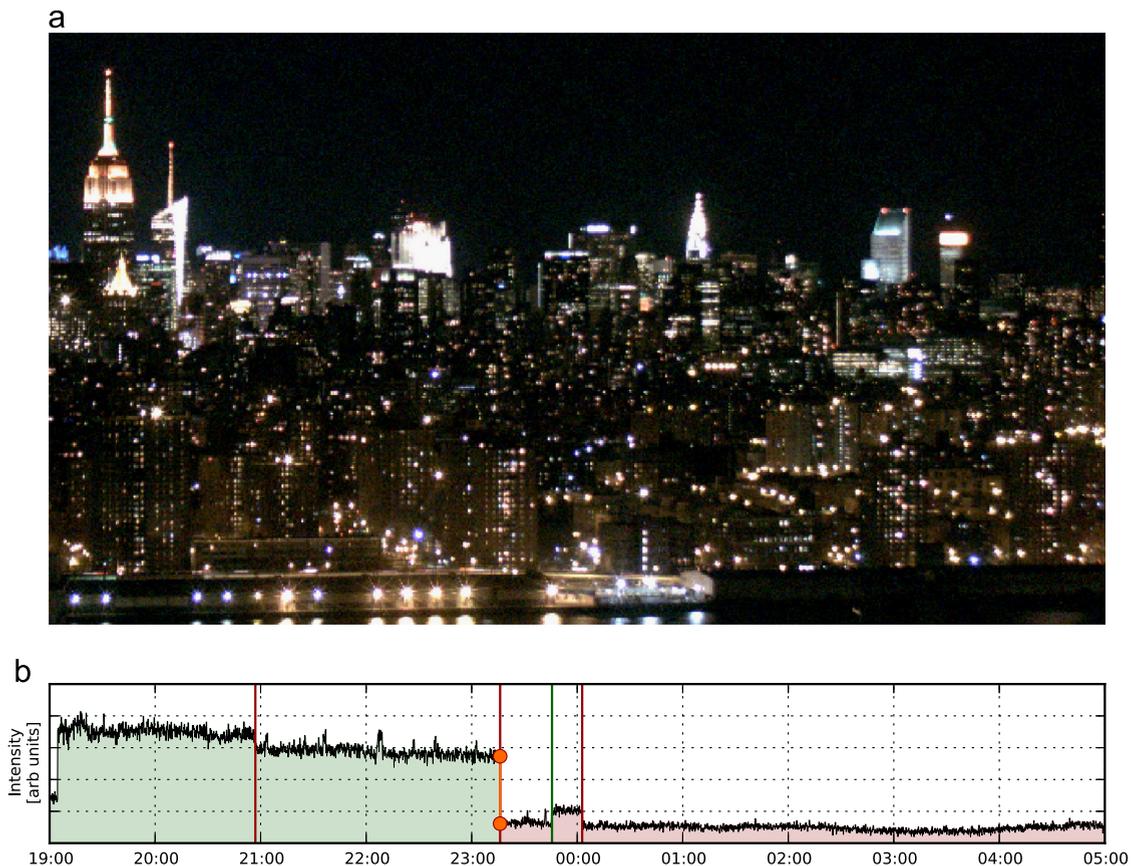
## 1. Introduction

A nighttime image of the Manhattan skyline (see Fig. 1a) contains thousands of artificial light sources, including windows, buildings, streets, vehicles, and billboards. While virtually all of these sources are stationary in position, many

are dynamic in time, changing in color and intensity throughout the night. The origins and timescales of variability are diverse, ranging from the 30 Hz flickers of a display screen to the few-Hz flickering induced by starting motors to discontinuities as shades are drawn or lights are turned on and off. Changing atmospheric conditions (for example the motions of clouds across the sky) also contribute variability. At low resolution, window lights in an urban night scene are analogous to variable stars on the night sky and so the techniques of observational astronomy

\* Corresponding author.

E-mail address: [greg.dobler@nyu.edu](mailto:greg.dobler@nyu.edu) (G. Dobler).



**Fig. 1.** The night scene viewed from the CUSP Urban Observatory. (a) This vantage point in downtown Brooklyn faces northward towards Manhattan. A time-lapse of the scene is available online at <https://www.youtube.com/watch?v=D3UKcoh6gig>. (b) A typical light curve for one of the sources in the scene. Vertical green (red) lines indicate on (off) transitions. The off transition corresponding to the largest change in average intensity, as measured before and after the transition (the  $t_{\text{off}}$  time), is shown in orange. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

can be applied to analyze the urban lightscape. In this paper, we demonstrate the utility of that approach in providing a new measure of urban activity and in revealing the dynamics of individual light sources.

While satellite observations of urban lights have been used to study city morphology, development, land use [12], energy consumption [14], and night lights [7], they are necessarily episodic and cannot probe dynamics on timescales shorter than a week. Observations from urban vantage points offer persistent coverage and an unchanging perspective, together with easy and low cost operations. Such images have been acquired for aesthetic purposes [13] but have not been analyzed for the scientific study of cities, with impacts from energy use and efficiency [8] to sleep patterns (which are a significant public health concern; see for example, [1,11]). *in situ* measurements with a comparable coverage, while perhaps more accurate, would be intrusive and would further entail the cost and operational difficulties of a large-scale sensor deployment.

In this work, visible light images were acquired from a rooftop in downtown Brooklyn, the first site of the Urban Observatory facility created by New York University's Center for Urban Science + Progress (CUSP). The northern

view across the East River covers the east side of lower and midtown Manhattan and offers a diversity of features, including the tops of the Empire State Building (at a distance of 6.1 km) and the Chrysler Building, major and minor building lights, and street and river lights. There are roughly 20,000 residential and commercial windows in the scene, and we estimate that some 100,000 people reside in the 4.4 km<sup>2</sup> covered by our images.

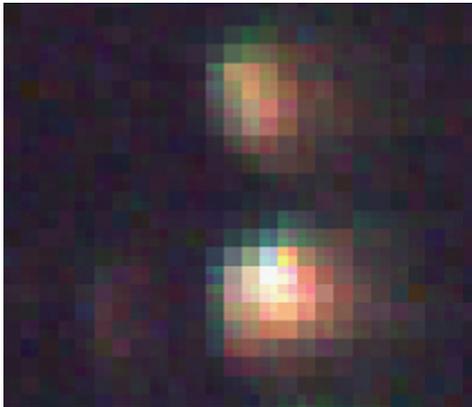
In Section 2 we describe our data acquisition and analysis pipeline which draws heavily from astronomical image processing and time series procedures, while in Section 3 we identify patterns in the light variability and the implications for aggregate versus individual behavior. We conclude in Section 4.

## 2. Methods

The images analyzed in this paper were acquired with a Point Grey Flea 3 8.8 megapixel camera (equipped with a 25 mm lens) every 10 s between 19:00 and 05:00 h on each of the 22 nights between October 26 and November 16, 2013. Daylight Saving Time ended during this period on November 3rd. Upon acquisition, each image (~25 MB in three-color raw format) was timestamped, encrypted, and

stored for analysis. A total of 70,560 images were acquired (equipment failure terminated observations at 00:00 on five nights), yielding a total data volume of  $\sim 4.5$  TB.

As we describe in detail below, images were processed in a two-stage pipeline: data reduction and data analysis. Our data reduction pipeline to extract light curves for each source begins by registering each image to a reference image to account for camera vibration and pointing drift. With good conditions frame-to-frame offsets were a few pixels (vertical drifts were of order 10 pixels over the 22 days). Registration failed if the shift exceeded 20 pixels in either direction and the correlation was not sufficiently close to unity, which could be caused by poor visibility due to weather or excessive vibration. Some 4147 rectangular window apertures were defined manually on a stacked image, comprising  $\sim 20\%$  of the total number visible ( $\sim 40\%$  of the near scene and  $\sim 20\%$  of the far scene). The intensity of each source within each frame was calculated by averaging the intensity of the pixels within the corresponding aperture in three colors (RGB). We refer to the time history of a source's intensity as its "light curve". The



**Fig. 2.** A zoom in of two of the closest window sources in our scene. Our privacy protections – which include limiting the maximum number of pixels per source – are enhanced by atmospheric effects which tend to blur the sources into amorphous blobs of light (see Fig. 1).

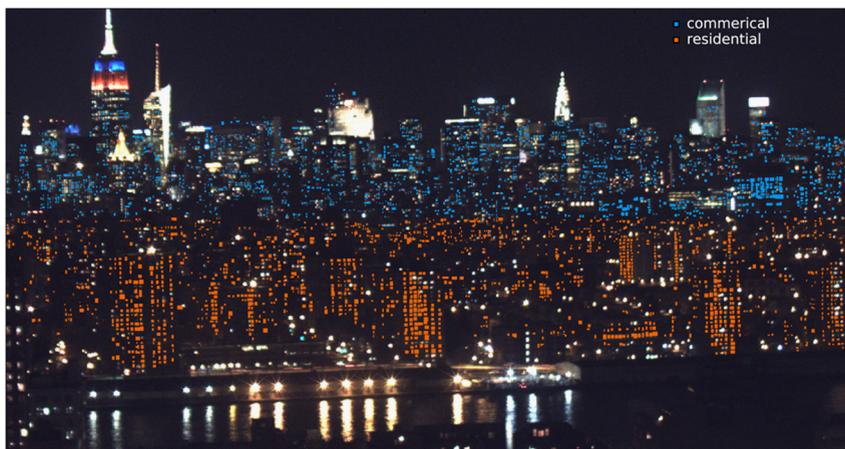
light curves were assembled with the time stamps of the individual images. Their accuracy is significantly higher than the 10 s cadence of our observations making that a negligible source of uncertainty.

The data analysis stage identifies on/off transitions in each light curve for each night and then from those, determines  $t_{\text{on/off}}$ , the on/off transition with the largest change in average intensity as measured before and after. For each light curve and for each night our algorithm first convolves the light intensity with a Gaussian filter of standard deviation 5 min, thereby reducing small-scale noise. On/off transitions are identified as peaks in the derivative of the convolved light curve. The situation is analogous to identifying edges in noisy images, so we use a variant of the well-known Canny edge detector [5]. This definition of on/off transitions minimizes effects due to drift and further cross checks were performed to guard against spurious on/off detections.

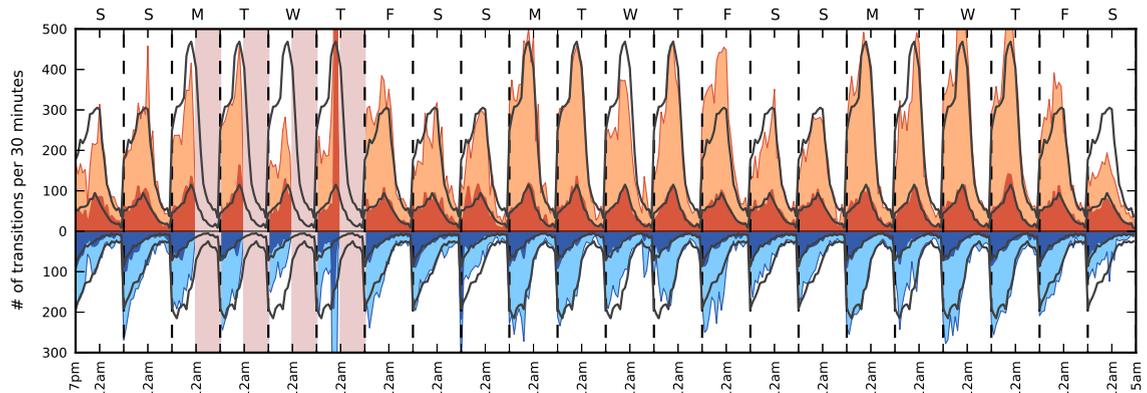
Protecting the privacy of those in our field of view has been paramount in our image acquisition and analysis and strict protocols have been observed toward that end. No more than a few pixels cover the closest sources in our scene, and atmospheric effects significantly blur the images further (see Fig. 2), ensuring that no personal detail is ever captured. In addition, all analyses have been performed at the aggregate level and any human inspection of an individual light curve has been done in ignorance of the precise location of that source within the scene.

### 2.1. Data reduction: image registration

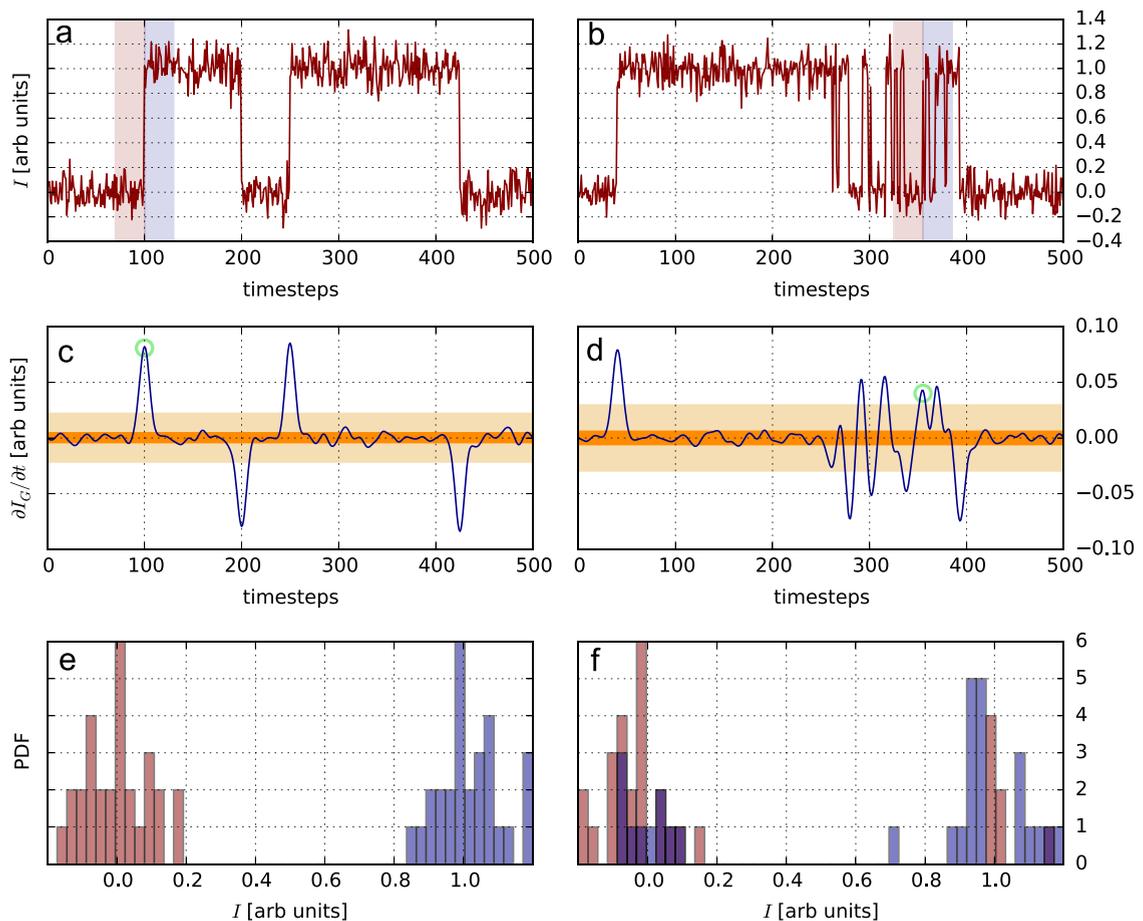
Wind-driven vibrations make the camera pointing time dependent, so that each frame must be registered to a reference image before extracting source intensities. The reference image was chosen by visual inspection to have many distinct features and be clear and in sharp focus. This image registration was performed by correlating a (mean subtracted and variance normalized)  $800 \times 800$  pixel sub-patch in each image with the corresponding patch in the reference image for various offsets in pixel row and



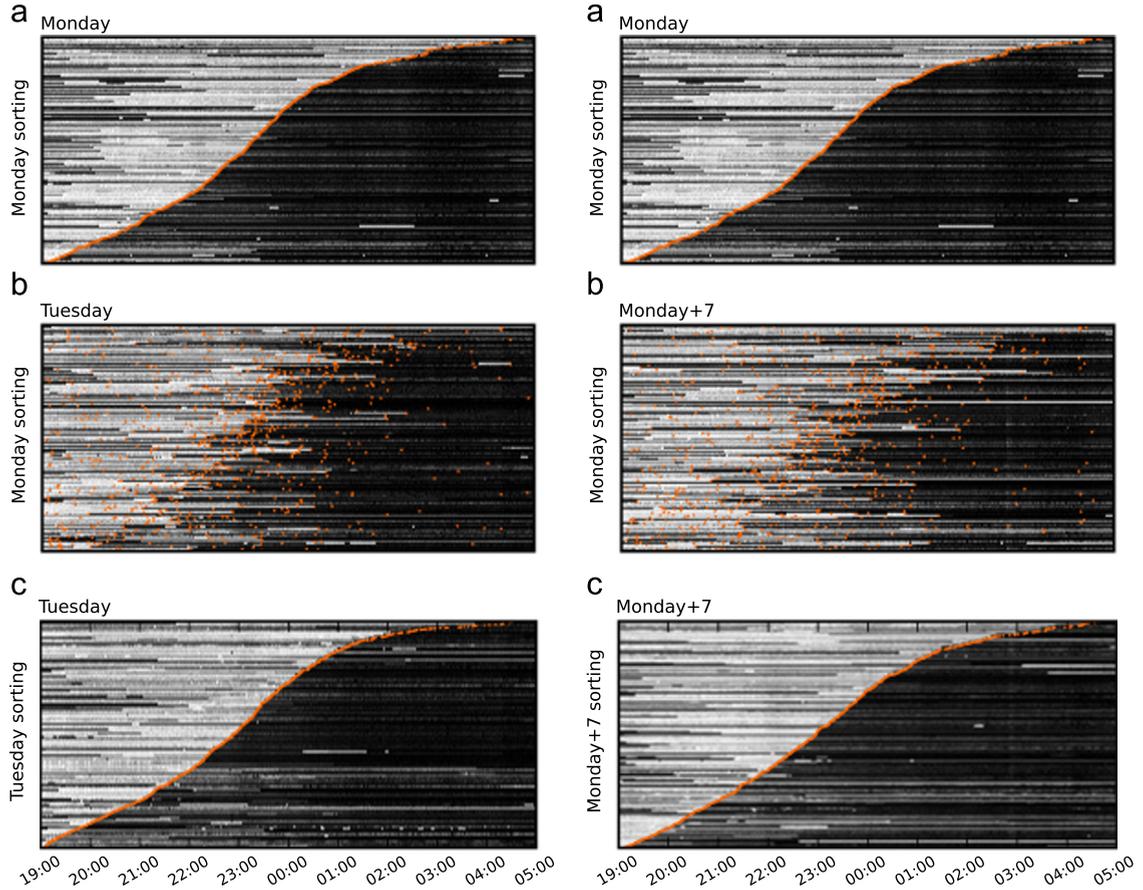
**Fig. 3.** The apertures defining the 4147 light sources in our scene. The apertures designated "commercial" and "residential" in our analysis (see Section 3.1) are shown in blue and orange respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



**Fig. 4.** The pulse of city lights curves over 22 nights. The number of transitions per 30 min is plotted for four transition types: all off transitions (light orange),  $t_{\text{off}}$  s (dark orange), all on transitions (light blue), and  $t_{\text{on}}$  s (dark blue). Pink regions denote missing data due to equipment malfunction and the solid curves are averages over the final 16 nights (M/T/W/T weekdays and F/S/S weekends). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



**Fig. 5.** An illustration of our robustness checks for the on/off transition detection algorithms using two example mock light curves. (a) A clean light curve with sharp, well-separated transitions. The only noise is the photometric noise of the observations. (b) A light curve with spurious transitions which can arise due to excessive camera vibrations or rapidly varying cloud conditions. (c) and (d) The derivative of the smoothed intensity for each light curve. In the absence of noise, peaks in this quantity represent locations of on/off transitions (for on/off spacing greater than the 5 min smoothing scale). Photometric noise introduces spurious, low-level peaks which we eliminate by performing 10 iterations of  $2\sigma$  outlier rejection. The final  $2\sigma$  band is shown in dark orange. Furthermore, we then only consider peaks which are  $10\sigma$  outliers from that clipped distribution (i.e., we only consider peaks outside the light orange band). (e) and (f) The light curve intensity distributions for the 5 minute interval before (red) and after (blue) the transitions circled in green in (c) and (d) (these time intervals are shown as shaded regions in (a) and (b)). For non-spurious transitions these distributions are well separated, while for spurious transitions there is significant overlap. We only consider transitions for which the separation is  $> 2\sigma$ . (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



**Fig. 6.** *Left:* the intensity as a function of time for primarily residential light curves for two nights and for different sortings. Orange dots denote  $t_{\text{off}}$  times of 1706 individual sources. (a) In the top panel light curves from Monday are sorted according to  $t_{\text{off}}$  times. (b) Tuesday light curves are plotted sorting according to Monday's  $t_{\text{off}}$  times and the pattern in (a) becomes strongly randomized. (c) Tuesday light curves are instead plotted sorting according to Tuesday  $t_{\text{off}}$  times and a curve similar to that observed on Monday re-emerges. *Right:* the same but comparing one Monday with the following Monday for the residential sample. Again, as with the day-to-day comparison, the week-to-week comparison shows that there is a repeatable pattern in aggregate, but that the individual sources do not exhibit strictly repeating behavior. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

column. Specifically, if  $I_i$  is the image to be registered,  $I_r$  is the reference image, and we take  $S_i$  and  $S_r$  to be the subpatch of the image and we reference respectively, we find  $(\Delta x, \Delta y) = (400 - x, 400 - y)$  where  $(x, y)$  are the column and the row which maximize the 2D convolution matrix

$$C = S_i * S_r' \quad (1)$$

where, for example,

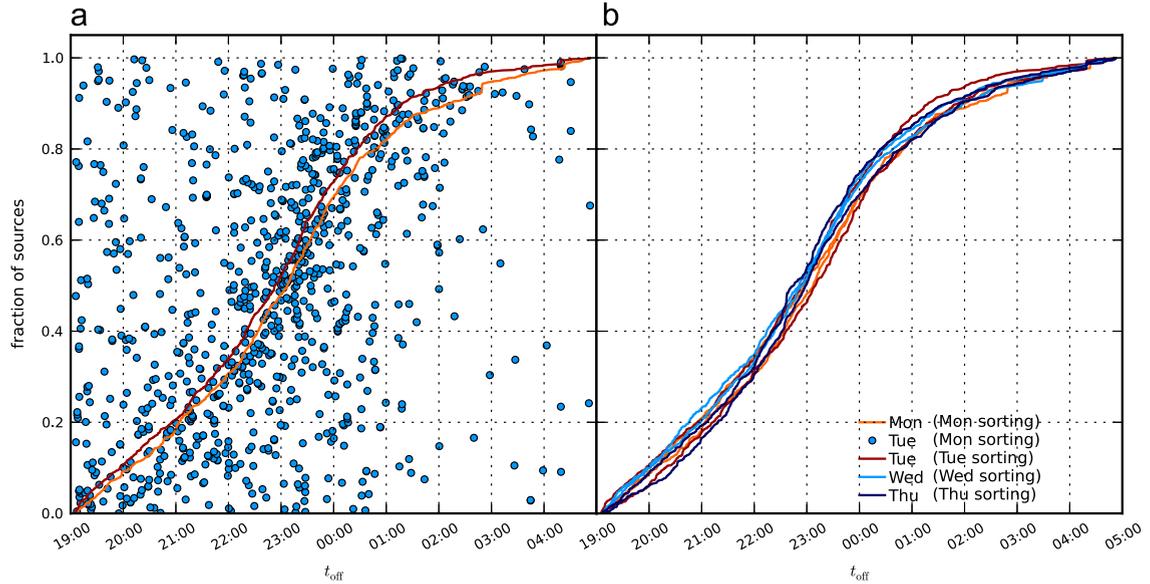
$$S_r' = \frac{S_r - \langle S_r \rangle}{\sqrt{\langle S_r^2 \rangle - \langle S_r \rangle^2}} \quad (2)$$

which sets the bounds of  $C$  to be  $[-1, 1]$ . Thus, the appropriate registration offset is determined by the pixel shifts that maximize correlation. With good conditions, frame-to-frame offsets were a few pixels (vertical drifts were  $\sim 10$  pixels over the 22 days). However, registration is deemed to fail (and the frame is not registered) if the shift exceeds 20 pixels in either direction and the correlation is not sufficiently close to unity. This might be caused by poor visibility due to weather or excessive vibrations. A total of 72,280 frames were successfully registered and there were only 79 failures. The final

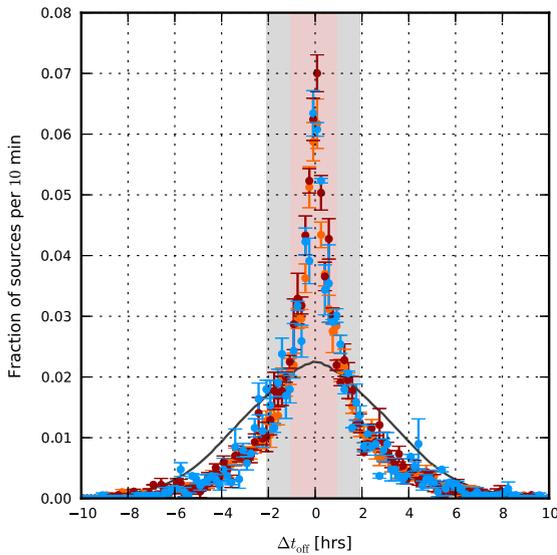
image size after registration is  $4056 \times 2120$  pixels, with a pixel depth of 8 bits in each of the RGB bands.

## 2.2. Data reduction: source selection

All registered night time images were stacked and, from the stacked image, apertures were placed around 4147 windows (of the  $\sim 20,000$  in our scene). Those rectangular apertures were assigned manually by selecting two diagonally opposed corners of the sources (upper left and lower right) which enclosed the light from that source. Aperture definition using the stacked image rather than a single image avoided biasing our sample toward lights that are only on at a specific time. The aperture selection included representation from all buildings, but, owing to low resolution at large distances as well as non-optimal viewing angles and the close spacing of some sources, the selection was sparser at larger distances. We estimate that  $\sim 40\%$  of sources in the near scene and  $\sim 20\%$  of sources in the far scene were identified for this study, which provides a sample size that is more than sufficient to yield statistically robust (see Section 3). The apertures are shown in Fig. 3.



**Fig. 7.** An overlay of the recurring  $t_{\text{off}}$  pattern in the residential sample. (a) The  $t_{\text{off}}$  curves for a given Monday with Monday sorting and Tuesday with Tuesday sorting for our residential sample. The blue points show the significant dispersion away from the pattern when the Monday sorting is used for the Tuesday points. (b) The  $t_{\text{off}}$  curve overlaid for 8 weekdays in our sample, showing the clearly repetitive nature of the aggregate pattern. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)



**Fig. 8.** The day-to-day variation in the change of  $t_{\text{off}}$ . The residential sample is shown in the histogram (binned in 10 min intervals) of the fraction of sources with a given change  $\Delta t_{\text{off}}$ . Comparison between 1 day (orange), 2 days (red), and 3 days (blue) are plotted. The gray line denotes a correlation histogram generated by pairs of independent draws from the distribution. The red (gray) shaded regions with boundaries located at  $\pm 1.0$  ( $\pm 2.0$ ) hours enclose 50% of the data (independent draw) distribution. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

### 2.3. Data reduction: source brightness and light curves

Brightnesses for each source were calculated in three bands by averaging RGB values of the pixels within that source's aperture. It is important to note that our brightnesses are not absolutely calibrated. However, for the

purposes of this paper, relative brightnesses of sources are sufficient as our results are based exclusively on feature identification in the light curves, which is independent of the absolute scale. The light curves were reduced to a single intensity  $I(t) = (R(t) + G(t) + B(t))/3$  (i. e., the mean of the three bands).

### 2.4. Data analysis: identifying on/off transitions

The time dependent behavior in our light curves can take several forms including noise fluctuations, slow zero point drifts, and sharp on/off transitions. Because the goal of our work is to assess patterns of activity, we have developed an algorithm (a modified version of a Canny edge detector in 1D) for identifying on/off transitions while suppressing both noise and drifts.

For each light curve  $I(t)$  (and for each night), our algorithm first convolves  $I(t)$  with a Gaussian filter of width 5 min (30 frames)

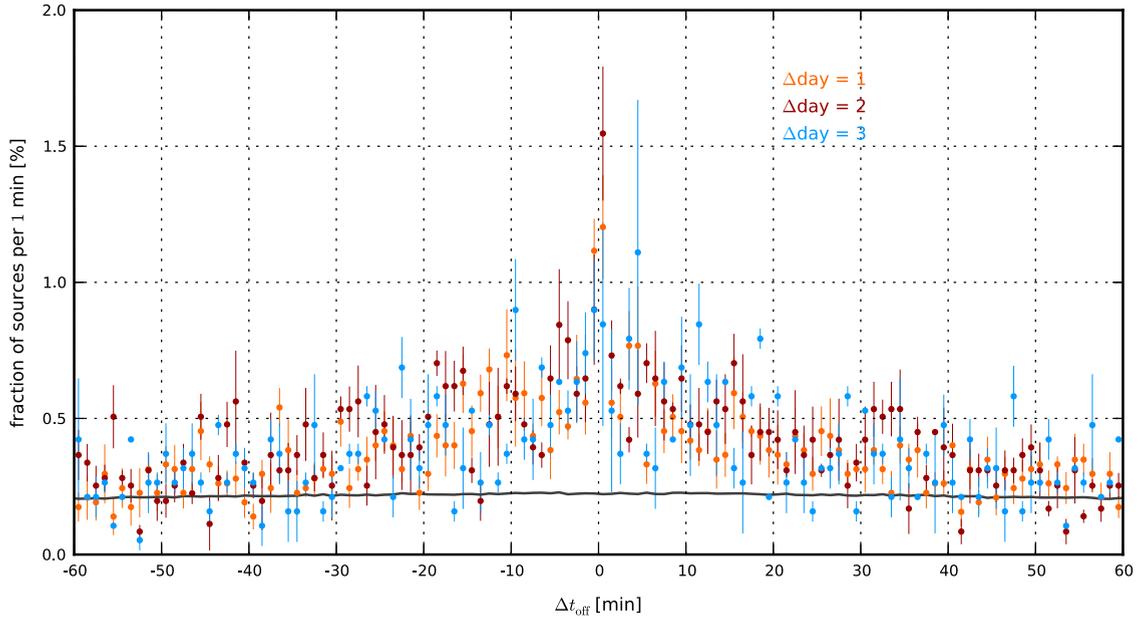
$$I_G = G_5 * I, \quad (3)$$

thereby reducing small scale noise. Edges (i.e., on/off transitions) are then identified as peaks in the derivative of this quantity so that

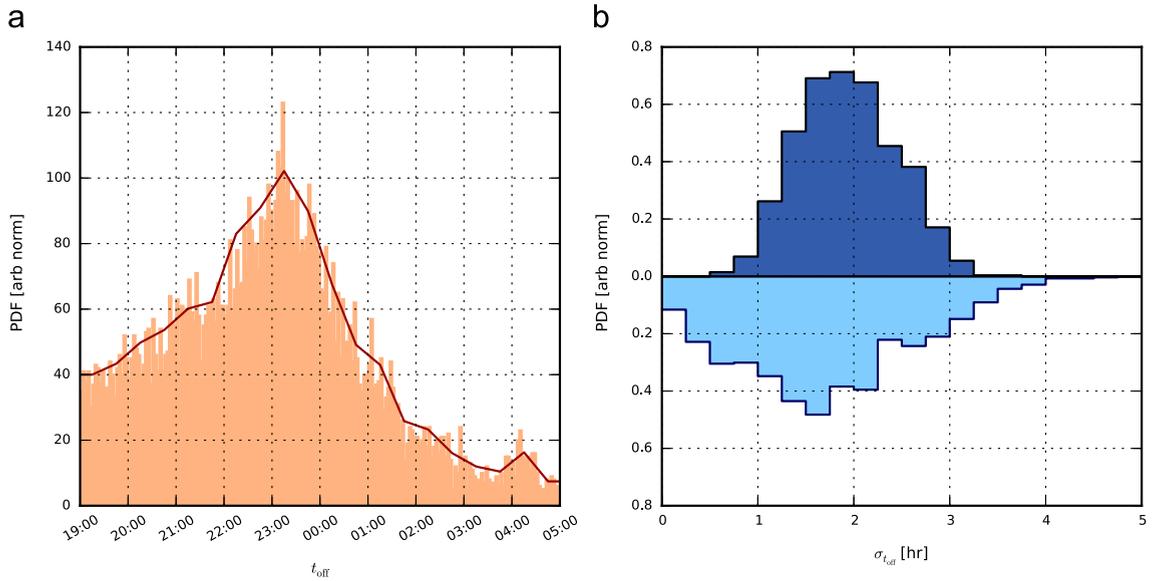
$$\ddot{I}_G(t = t_{k,\text{on/off}}) = \frac{\partial^2 I_G}{\partial t^2} = 0, \quad (4)$$

where  $k$  indexes the transitions in a given light curve (it is typically less than 5). By concentrating on peaks in the derivative of  $I_G$ , this definition of  $t_{k,\text{on/off}}$  minimizes effects due to drift.

We have two tests to protect against spurious transitions due to noise. These are illustrated in Fig. 5. In our first test, we restrict our transitions in Eq. (4) to only those peaks in  $\dot{I}_G$  which are  $10\sigma$  outliers by performing recursive



**Fig. 9.** The same as Fig. 8 but with a binning resolution of 1 min. The sources which have  $t_{\text{off}}$  times within a minute from one day to the next (likely lights on timers) comprise roughly  $\sim 2.5\%$  of our sample.



**Fig. 10.** A comparison of the individual  $t_{\text{off}}$  behavior with an independent and identically distributed random draw. *Left:* the average  $t_{\text{off}}$  distribution from our residential observational sample (red) and an IID draw of 1100 sources for 8 days (orange) from that parent distribution. *Right:* for a given source, the variance in  $t_{\text{off}}$  is calculated for both the observations (light blue) and IID draw (dark blue). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

$2\sigma$  outlier rejection for 10 iterations. Then, we look for 10  $\sigma'$  outliers in  $l_C$  where  $\sigma'$  is the standard deviation of the data points after outlier rejection. For our second test, we check that, for each transition, the mean of the light curve for 5 min before the transition is  $2\sigma$  away from the mean of the light curve for 5 min after the transition where  $\sigma$  here is the maximum noise in the light curve during those two time intervals. This test is particularly effective in poor weather conditions where very noisy light curves can

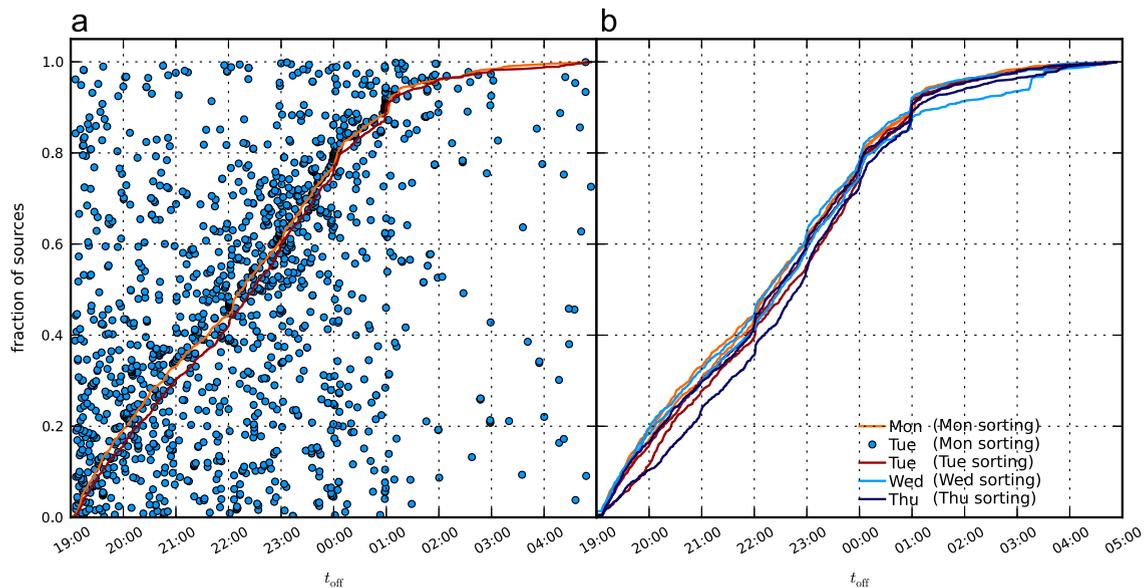
induce many spurious edge detections separated closely in time.

Lastly, we formally define the “big off” time for each light curve as

$$t_{\text{off}} = \arg \max_k [\delta(t_{k,\text{off}})], \tag{5}$$

where

$$\delta(t_{k,\text{off}}) = \langle I(t < t_{k,\text{off}}) \rangle - \langle I(t \geq (t_{k,\text{off}})) \rangle. \tag{6}$$



**Fig. 11.** The same as Fig. 8 but for the commercial sample of lights on weekdays. As in the residential sample, there is a repeated pattern night-to-night and week-to-week with more random individual activity. The most notable difference between the residential and commercial samples is the lack of inflection in the  $t_{\text{off}}$  curves, indicating a more random distribution of  $t_{\text{off}}$  prior to midnight, by which time most commercial sources have experienced their  $t_{\text{off}}$ .

That is,  $t_{\text{off}}$  is the time of that transition having the largest difference between the mean value of the intensity before and after. We define  $t_{\text{on}}$  analogously.

### 3. Results

A typical light curve is shown in Fig. 1b. Small high-frequency noise fluctuations are interrupted by clear discontinuities (transitions) that signal activity. While we have not attempted to analyze the causes of any particular transition, we note that large discontinuities could be caused by turning a light on/off or by opening/closing a shade or drape, while small discontinuities might be associated with a small light (e.g., desk lamp) or a translucent curtain or an interior room light. As described in Section 2, transitions where the intensity decreases (increases) are referred to as “off” (“on”) transitions, and we denote by  $t_{\text{off}}$  the time of that “off” each night having the largest decrease in average intensity before and after (the orange vertical line in Fig. 1b; similarly the largest “on” occurs at  $t_{\text{on}}$ ). This transition detection drastically reduces the data volume from several TB of images to a few transition times per night for each source observed. The aggregate of these transitions over many nights is one dimension of the pulse of the city. This is evident in Fig. 4 which shows four approximately repeating patterns obtained from the on and off transitions for our complete sample. Given the very different use cases for commercial versus residential buildings in urban environments and the anticipated difference in weekdays versus weekends, residential/commercial and weekday/weekend subdivisions of our sample are discussed below.

#### 3.1. Light patterns from residential sources on weekdays

The lower half of our scene contains largely residential buildings, while those in the upper half are largely

commercial buildings in midtown Manhattan. As we expect residential and commercial buildings to have different activity patterns, we have isolated a primarily residential sample of 1706 sources comprising only those in the lower half of the scene. Those residential light curves are plotted in grey scale through a Monday → Tuesday night in Fig. 6a; the sources have been sorted according to their  $t_{\text{off}}$  times (orange points). The resulting curve is the cumulative distribution of  $t_{\text{off}}$ , which shows an inflection point near 23:00 corresponding to a peak rate of  $t_{\text{off}}$  events. In Fig. 6b the residential light curves for the following night (Tuesday → Wednesday) are shown in the same sorting as that of Fig. 6a (i.e., Monday’s sorting). If each source were to turn off at roughly the same time each night, Fig. 6b would show roughly the same cumulative distribution; instead, the  $t_{\text{off}}$  times become largely disordered. However, Fig. 6c shows that if Tuesday’s light curves are instead sorted by their own  $t_{\text{off}}$  values, a pattern very similar to Monday’s re-emerges.

It is reasonable to suspect that, while we have shown that individual light activity does not strictly repeat day-to-day, there may be a weekly coherence of individual sources. In the right hand panels of Fig. 6 we show the behavior of Monday → Monday + 1 week. As with the day-to-day comparison, the aggregate pattern recurs from one Monday to the next, but individual behavior is strongly decorrelated. In fact (as shown in Fig. 7) the macro pattern persists for all full weekday nights in our sample despite the more random micro behavior. In other words, while individual sources do not habitually turn off at the same time each night, the aggregate activity is quite repeatable.

Some insight into the single-source behavior underlying these macroscopic regularities can be obtained by asking “By how much does  $t_{\text{off}}$  vary for individual sources

between nights separated by one, two, or three days?” Fig. 8 shows a histogram of such variations over all of the residential sources. The overlap of the distributions for separations of 1, 2, and 3 days indicates randomization after only 1 day. Nevertheless, a comparison of these distributions with a random draw from the parent population of  $t_{\text{off}}$ s (gray line) shows stronger night-to-night correlation than random. However, a full 50% of our residential sample has  $|\Delta t_{\text{off}}| > 1.0$  h, again indicating that a substantial fraction of our sample do not have regular  $t_{\text{off}}$  times from one night to the next. We also note that  $\sim 2.5\%$  of our residential sample repeats nightly with a variation of less than 1 min as shown in Fig. 9, suggesting the presence of timers.

Further evidence of non-trivial single-source dynamics can be seen by comparing (for each source) the variance in  $t_{\text{off}}$  among the 8 weeknights in our sample for which we

had a full 10 h of observations (see Fig. 4) with that of a random draw of octuples from the parent  $t_{\text{off}}$  distribution. Fig. 10 shows the probability distribution function for  $t_{\text{off}}$  averaged over 8 weekdays of our residential sample and an independent and identically distributed (IID) random variable draw of 1100 samples of 8 points (i.e., 8 days) from that parent distribution. The distribution of variances in the  $t_{\text{off}}$ s over the 8 nights for the residential sources in our observational sample is markedly different than for the IID draw. The broader distribution in the data indicates the difference in source dynamics compared to a random draw from the aggregate parent population. Specifically, there are some sources which have more repeatable  $t_{\text{off}}$ s (within about 1 h) while others have much more random behavior than what would be expected from a purely random draw of  $t_{\text{off}}$ s from the average weekday distribution.

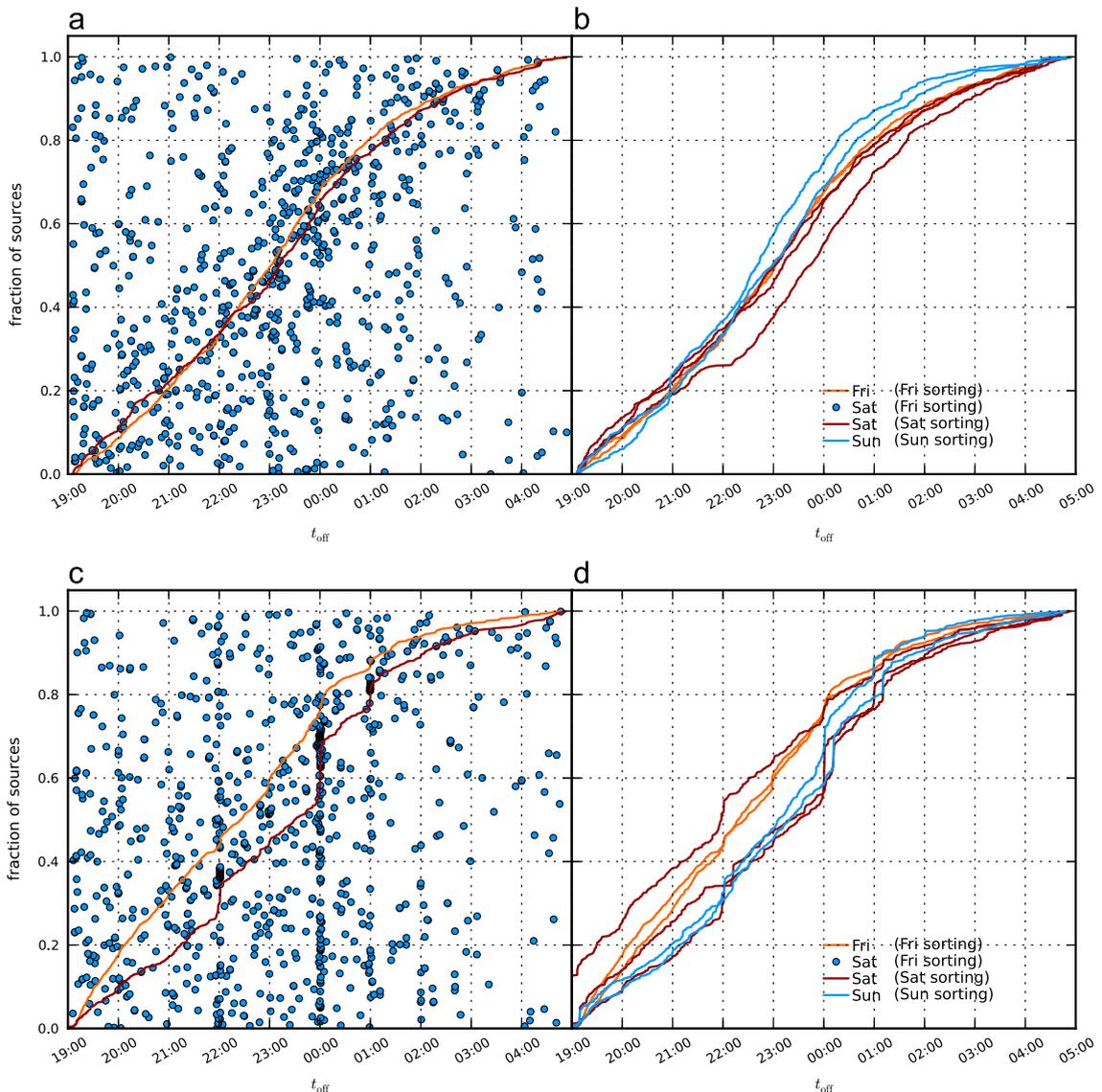


Fig. 12. Top (bottom): the same as Fig. 7 but for the residential (commercial) sample of lights on weekends.

### 3.2. Light patterns from commercial sources on weekdays

The overlay of the aggregate  $t_{\text{off}}$  pattern for the primarily commercial scene is plotted in Fig. 11. It also exhibits repeating, albeit distinct, aggregate behaviors. However, before about 1:00 am we do not observe the significant curvature seen in the residential sample, indicating that there is not necessarily a characteristic “off” time for commercial lights, and their transition activity before midnight is more random. There are however multiple lights (including lights on timers, whole floors, splash from spotlights, etc.) that turn off on the hour in the commercial sample.

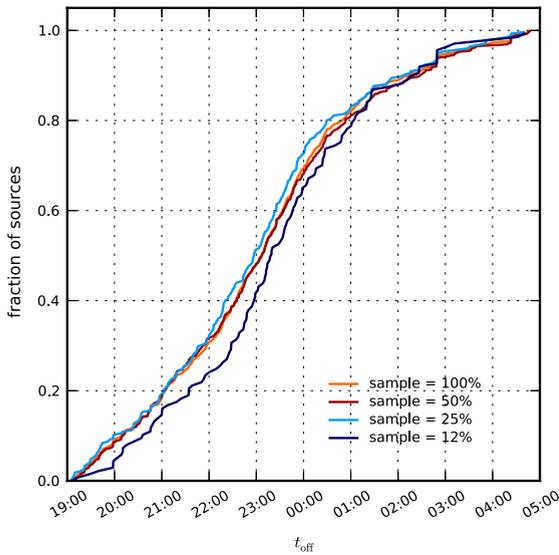


Fig. 13. The  $t_{\text{off}}$  pattern for random subsamplings of the data.

### 3.3. Light patterns on weekends

An overlay of the recurring  $t_{\text{off}}$  pattern on weekends for both residential and commercial activity is plotted in Fig. 12. Each exhibits a pattern distinct from that seen on weekdays. Again, as for weekdays the aggregate pattern emerges, but there is more scatter on the weekends when comparing successive days (note, the flat behavior on one Saturday between about 8:30 pm and 9:15 pm is due to missing data). Interestingly, the Saturday with Friday sorting points are less tightly clustered around the Friday with Friday ordering curve than the Monday → Tuesday comparison shown in Fig. 7, indicating that the weekend behavior is in fact more random than the weekday behavior.

### 3.4. Robustness to sub-sampling: residential and commercial

To test the robustness of our results (namely repeatable aggregate behavior with more random individual behavior) we perform subsamplings of our data over various dimensions in the analysis. First, the  $t_{\text{off}}$  curve for various random subsamplings of the residential sample on a Monday night is shown in Fig. 13. In each case, we randomly subsample some fraction of the total number of residential light sources and reorder according to  $t_{\text{off}}$  for that subsample. The fact that the shape of the  $t_{\text{off}}$  curve is robust to this subsampling (even down to 12% of the sample) indicates that our parent sample is sufficiently large. Fig. 14 shows the  $t_{\text{off}}$  pattern for only large amplitude transitions. Specifically, the sample in panel (a) (b) represents sources which the brightness after transition is  $\leq 70\%$  ( $50\%$ ) of the brightness before the transitions. Thus, our main results (repeatable pattern in aggregate but less repeatable on an individual basis) are robust to restricting ourselves to only the largest  $t_{\text{off}}$  transitions. Lastly, in

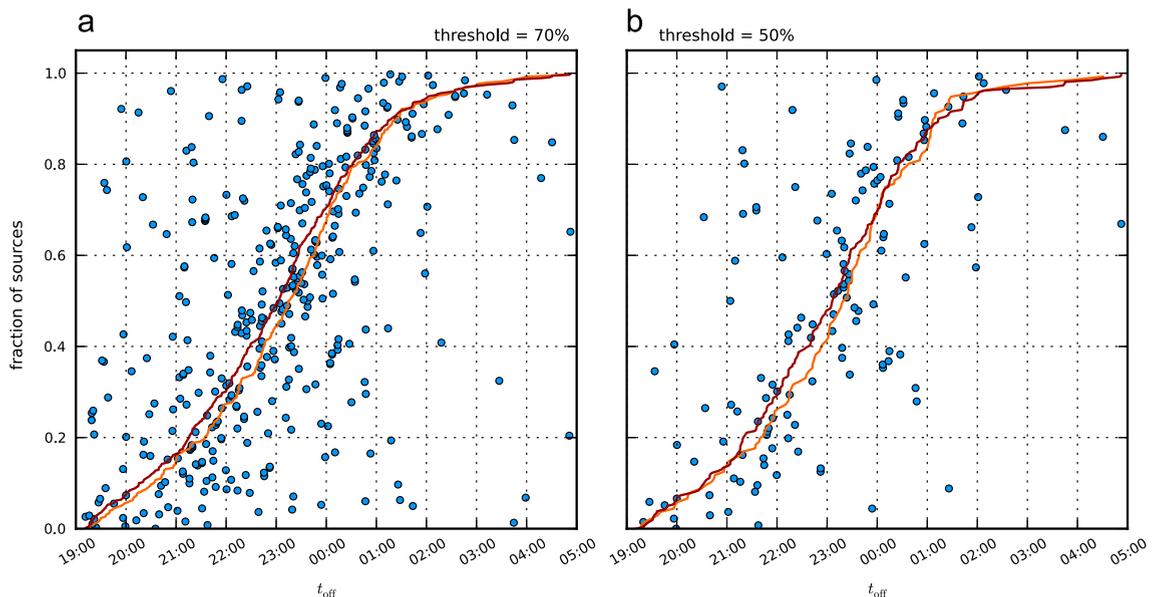
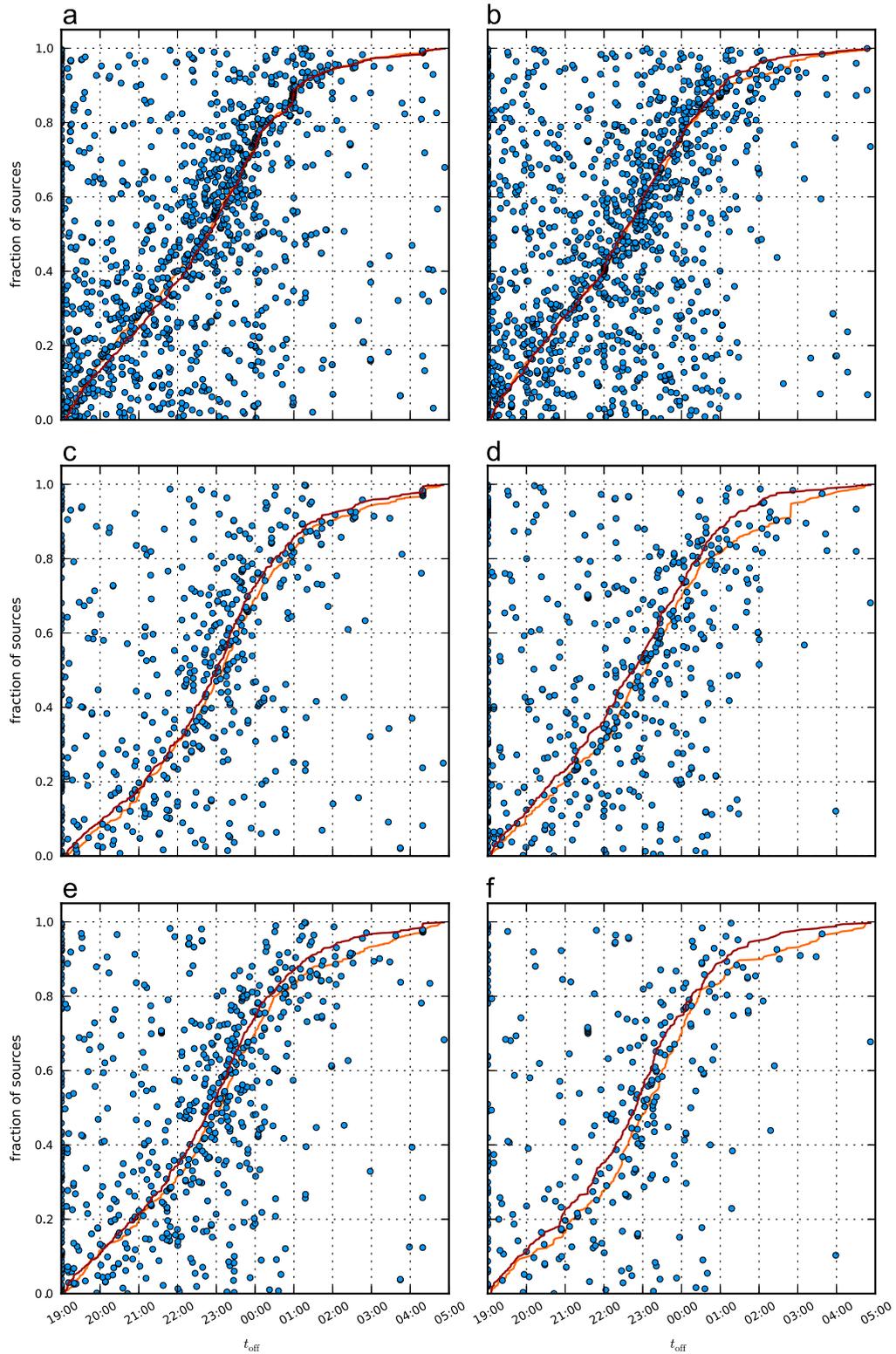


Fig. 14. The same as Fig. 7a but for subsamples of  $t_{\text{off}}$  that are restricted to different amplitude thresholds.



**Fig. 15.** The same as Fig. 7a but for different spatial subdivisions of sources in our scene: ((a) and (b), left and right half of the whole sample; (c) and (d), left and right half of our residential sample; (e) and (f), different definitions of the dividing line of residential vs. commercial. In each case, our results hold for different spatial subsamplings.

Fig. 15 we show that various spatial subsamplings also yield similar macro/micro behavior in the  $t_{\text{off}}$  distributions.

#### 4. Conclusions

We have described a fundamentally new methodology for quantifying the “pulse” of a city via time-dependent changes in the urban lightscape. Our use of astronomical techniques to analyze the dynamics of city lights applies physics to the new realm of urban science and human behavior, complementing the use of statistical physics for scaling and modeling of urban growth [10,2]. The latter studies use existing data sources and examine slow phenomena (e.g., urban sprawl) while our work focuses on rapidly varying urban signals.

Furthermore, we have demonstrated that an astronomy-inspired analysis of persistent synoptic imagery of an urban scene leads to insights into aggregate light activity. In particular, we have discovered a repeatable pattern of behavior characterized by the aggregate on/off transition times of lights in the scene which recurs night-to-night and week-to-week. Despite the strong repeatability of that pattern, the individual sources which comprise it do *not* strictly repeat night-to-night or week-to-week. Specifically, a full 50% of our sample decorrelates by more than an hour after only one night. This decorrelation does not significantly increase when comparing nights separated by 2, 3, or 7 days. The same qualitative behavior is seen for both residential and commercial light sources and for both weekdays and weekends.

Interestingly, however, the individual source dynamics of our urban lights sample is more complex than a simple random draw from the aggregate parent distribution of off transitions which repeats night-to-night. We find that the distribution of nightly correlations in our sample of residential sources is broader than that expected from an independent and identically distributed draw from the repeating pattern.

Our results represent the regularity and variability of one component of the pulse of the city. A much more random, but temporally correlated, individual behavior has been found to underlie that regularity. The method we have described in this paper represents a first step in a new approach to studying urban functioning. Correlations of the kind of data we have presented with such variables as demographics, energy use, season, income, and meteorological conditions are of clear interest and will be the focus of our future work. Disturbances in these patterns by light pollution and noise pollution will

give insights into the public health affects on the circadian rhythms of city dwellers while building level aggregation will inform measurements of occupancy (a strong correlate for energy consumption). Further, beyond basic urban science, applications include emergency response, environmental monitoring, and urban operations, and the CUSP Urban Observatory will push this methodology to other observational modalities including broadband infrared and hyperspectral observations.

#### Acknowledgments

This work was partially supported by a grant from the Alfred P. Sloan Foundation grant # 2013-10-37.

#### References

- [1] A. Abbott, *Restless nights, listless days*, *Nature* 425 (2003) 896.
- [2] C. Andersson, K. Lindgren, S. Rasmussen, R. White, Urban growth simulation from “first principles”, *Phys. Rev. E* 66 (August (2)) (2002) 026204, <http://dx.doi.org/10.1103/PhysRevE.66.026204>.
- [3] L. Bettencourt, *The origin of scaling in cities*, *Science* 340 (2013) 1438.
- [4] L.M.A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G.B. West. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci.* 104 (17) (2007), 7301–7306, <http://www.pnas.org/content/104/17/7301.abstract>.
- [5] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* (1986), <http://dx.doi.org/10.1109/TPAMI.1986.4767851>.
- [6] N. Ferreira, J. Poco, H.T. Vo, J. Freire, C.T. Silva, Visual exploration of big spatio-temporal urban data: a study of New York city taxi trips, *IEEE Trans. Vis. Comput. Graph.* 19 (12) (2013) 2149–2158.
- [7] J. Hale, G. Davies, A.J. Fairbrass, T.J. Matthews, C.D.F. Rogers, et al., Mapping lightscape: spatial patterning of artificial lighting in an urban landscape, *PLoS ONE* 8 (5) (2013) e61460.
- [8] C.E. Kontokosta. Local Law 84 Energy Benchmarking Data. Report to the New York City Mayor's Office of Long-Term Planning and Sustainability, 2012.
- [9] T. Louail. From Mobile Phone Data to the Spatial Structure of Cities, [arXiv:1401.4540](https://arxiv.org/abs/1401.4540), 2013.
- [10] H.A. Makse, J. Andrade, M. Batty, S. Havlin, H.E. Stanley, Modeling urban growth patterns with correlated percolation, *Phys. Rev. E* 58 (December (6)) (1998) 7054–7062, <http://dx.doi.org/10.1103/PhysRevE.58.7054>.
- [11] T. Roenneberg, K.V. Allebrandt, M. Meroow, C. Vetter, Social jetlag and obesity, *Curr. Biol.* 22 (10) (2012) 939–943. URL <http://www.sciencedirect.com/science/article/pii/S0960982212003259>.
- [12] P. Sutton, A scale-adjusted measure of “urban sprawl” using nighttime satellite imagery, *Remote Sens. Environ.* 86 (2003) 353–369.
- [13] A. Warhol. *Empire*, 1964.
- [14] J. Wu, Z. Wang, W. Li, J. Peng, Exploring factors affecting the relationship between light consumption and GDP based on dmisp/ols nighttime satellite imagery, *Remote Sens. Environ.* 134 (0) (2013) 111–119. URL <http://www.sciencedirect.com/science/article/pii/S0034425713000734>.