Computers, Environment and Urban Systems 68 (2018) 133–141

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Seasonal population estimates based on night-time lights

Demetris Stathakis*, Pavlos Baltas

University of Thessaly, Dept. Planning and Regional Development, Athinon 38334, Greece

ARTICLE INFO

Keywords:
Seasonal population
Human geography
Earth observation
Night-lights
VIIRS

ABSTRACTS

The objective of this paper is to present a method for estimating seasonally specific ambient population counts. The central assumption is that the variation in observed night-lights is a valid proxy for ambient population. Island populations are used for validation, where it is possible to derive estimates of ambient population from national statistics. The method is then applied to the whole of Greece. The validation shows a strong correlation amongst night-lights derived estimates and the reference dataset. Based on the proposed method, national maps are produced showing the month when seasonality is in its peak, the peak value during that month and the overall length of the season, in terms of how many months exceed a certain threshold. Different seasonality patterns are revealed. An advantage of the proposed method, compared to other contemporary approaches, is that it is based on public domain, global data.

1. Introduction

Population counts and estimates are normally available from censuses and registers. Census residential population is counts of people at their permanent residencies (a.k.a. "night-time" population). Ambient population, as opposed to residential population, is the total population present at each particular location at a given time (Amaral, Camara, Miguel Vieira Monteiro, Alberto Quintanilha, & Elvidge, 2005; Sutton, Elvidge, & Obremski, 2003). The main benefit of ambient population is that it can be used to track the movement of people within the day (a.k.a "day-time" population). Ambient population shows for example the concentration of people in shopping centers during the day and in stadiums during the weekends, both void of resident population. Censuses are typically performed once per decade. But people spend time in places other than their residence, generating a demographic concept known as the seasonal population. People move in space for multiple reasons, including commuting to work and taking holidays. Depending of the length of stay, different seasonal population groups are formed, with quite distinct characteristics. A first group refers to tourists traveling to places with some kind of touristic interest. Tourists that only make short stays are also known as visitors. A second group includes seasonal workers attached to seasonal jobs (in tourism, agriculture, construction etc.). Third is the group that includes second-house owners. They are typically expected to spend more days than tourists at their second-house location. Fourth, migrants, registered or unregistered, that move in space for several reasons including refugees escaping conflicts as well as people moving due to socioeconomic factors.

Seasonal population, typically exhibits a peak time at each place. Seasonality peak-time depends on what is actually attracting the additional, non-resident, people. For tourists it can be summer holidays, in coastal areas, or winter holidays, in mountainous regions. It may be religious events triggering the movement of people throughout the year (Roman & Stokes, 2015). In any case, time-specific population distributions within a year are hard to estimate by conventional means. Custom surveys are inevitably limited in scope and costly, rendering repetition infrequent. Proxy variables such as water and electricity consumption can also be used to estimate seasonal population. The reliability of this approach is however reduced by the fact that (i) consumption data are available at two or three month intervals (ii) billing is sometimes based on estimated rather than actual consumption, for long periods of time before an actual measurement is made (iii) ephemeral events are present in the data (measurement errors, leaks etc.) (iv) per capita consumption of water and electricity is not constant throughout the year, consumption patterns change during the summer due to swimming pools, increased irrigation needs, excessive use of air conditioning etc.

Occupancy of tourist accommodation establishments is also frequently used as a metric for estimating the additional present population. The recorded data include percentage of beds occupied in the establishments. However, this data is affected by the shadow economy (illegally operating establishments) and by tax evasion (legally operating establishments not declaring all stays). Recently, occupancy data are becoming even more unreliable due to the rapid takeover of disruptive technologies such as the AirBnB platform (Airbnb, 2017) which

* Corresponding author.
E-mail addresses: dstath@uth.gr (D. Stathakis), pbatlas@uth.gr (P. Baltas).

https://doi.org/10.1016/j.compenvurbsys.2017.12.001
Received 18 June 2017; Received in revised form 5 September 2017; Accepted 3 December 2017
Available online 16 December 2017
0198-9715/

currently escapes national statistics. It has been suggested that Airbnb already provides a viable alternative for certain traditional types of overnight accommodation (Zervas, Proserpio, & Byers, 2016).

The phenology of population and the ability to capture and anticipate its dynamics has crucial implications in several domains. For example, seasonal population is useful in risk management in order to better estimate population at risk at each time (Smith et al., 2015). It is also important for improving energy demand forecasting (Roman & Stokes, 2015) and in epidemiology to monitor disease outbreaks (Bharti et al., 2011). In planning, both the amount and the type of seasonal population is essential because it substantially alters the demand for services. The demographic and socioeconomic profile of residents vs. the several non-residential groups is typically very different. Imposing a different strain on services and infrastructure.

Seasonal population is also one of the primary factors in determining the carrying capacity of a particular place. Carrying capacity is defined as the point where a place becomes insufficient to meet, without degradation, the needs of both resident and seasonal population due to natural or anthropogenic (infrastructure) constrains (Coccossis & Parpairis, 2000). Respecting the capacity of the local system is important both in the context of sustainable development and for maintaining the attraction and competitiveness of touristic destinations (Coccossis & Mexa, 2004).

Recent efforts to estimate seasonal population have mainly focused on mobile phone-call records (Deville et al., 2014, Erbach-Schoenberg, Alegana, Sorichetta, Lianard, et al., 2016, Hanaoka, 2016, Ratti, Frenchman, Pulselli, & Williams, 2006, Reades, Calabrese, & Ratti, 2009, Wesolowski et al., 2015, Wilson, Erbach-Schoenberg, Albert, Power, et al., 2016, Yang, Fang, & Xu, 2016). Essentially, phone-call records are treated as big-data. While this dataset provides powerful insights in population movements, it is limited by the fact that phone call data are not public domain. Therefore usage is restricted to those having access to the data. Roman and Stokes (2015) recently exploited a different data source, daily satellite night-light images, revealing seasonal movement at fine intervals of space and time. However, the night-light averages used for periods of less than a month are currently also not public domain.

In general, the strong correlation of night-lights with population has been proven in numerous studies. Elvidge, Hsu, Baugh, & Gosh, 2014 observed a strong correlation between night-lights and population in most countries of the globe. Stathakis (2016) noted a very strong correlation between night-lights and resident population for Greece in specific. Amaral et al. (2005) noted a strong correlation between night-lights and urban population in Brazil. Other than demography, previous studies used night-lights in numerous domains such as to assess the economic performance of areas (Ma, Zhou, Pei, Haynie, & Fan, 2012; Triantakonstantis & Stathakis, 2014, chap. 18; Stathakis, Tselios, & Faraslis, 2015), urbanization (Stathakis, 2015; Zhang & Seto, 2011; Zhang & Seto, 2013) health-related topics (Kloog, Haim, Stevens, & Portnov, 2009) and conflicts (Li & Li, 2014).

Our objective is to propose a method to estimate seasonal population based on global, publicly available data. The proposed approach is based on monthly composites of night-light satellite images that only recently became available. The dataset is described in detail in the next section. The method to use it as a proxy for monthly population estimates is then introduced. The main novelty of this approach is that the strong correlation of night-lights with population is exploited to derive monthly rather than annual estimates, based on a new method developed to efficiently process the satellite data. We believe that the proposed approach can fill the current data-gap with respect to monthly population estimates and trigger practical applications in this newly available human-geography time-scale.

2. Data and study are

The main information source to base the proposed estimates of seasonal population are the night-lights as recorded by the Visible

Table 1
Comparison of main characteristics of OLS and VIIRS sensor.

	DSMP/OLS	SUOMI/VIIRS
Spectral bands commonly used	1 panchromatic	1 panchromatic
Radiometric resolution	6 bits, values in [0, 63]	12 bits, values in [0, 4096]
Temporal resolution of processed products	1 year	1 month
Spatial resolution of processed products	1 km	0.75 km
Suitable for time series analysis	Only after inter-calibration	Yes, (ephemeral lights not removed)
Time-series	1992–2013	2012 - today uncorrected 2014 – today stray-light corrected

Infrared Imaging Radiometer Suite (VIIRS), Day/Night Band (DNB) sensor on-board the SUOMI satellites (Mills, Weiss, & Liang, 2013). VIIRS data are average radiance composite images having excluded data impacted by cloud-cover. VIIRS is the successor of the previously used Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) sensor. The DSMP/OLS series was discontinued in 2013 (Elvidge et al., 2014). The main advantages of VIIRS compared to the OLS are shown on Table 1 and explained in the remaining of this section.

In a nutshell, the improvements by order of importance are (i) spectral (ii) temporal and (iii) spatial resolutions. The major advancement is the improved spectral resolution which now is sufficient to overcome the saturation problem that significantly affected OLS (Elvidge et al., 2014). The term saturation describes the situation when the sensor records the maxim permissible value (DN = 63) while the observed value is actually higher. Due to the initial purpose of operation (detection of clouds in the night), OLS was typically operated in a high-gain setting i.e. by amplifying the incoming signal. A side effect of this amplification is that the output signal is saturated above brightly lit urban centers. Consequently, variations within urban centers cannot be detected. The second major improvement is temporal resolution. Whereas OLS data are distributed as annual rasters (a product known as ‘stable lights’), VIIRS data are distributed as monthly composites. This improvement in temporal resolution opens up a new spectrum of applications, including the one presented here. The last improvement is in spatial resolution. The pixel size became smaller. While this improvement is enough to better distinguish features, its magnitude is not enough to trigger new applications.

Nevertheless, VIIRS data has two major limitations. First, its time-series is currently significantly shorter compared to that of OLS. In fact, the time-series of VIIRS is currently so short it can only be used for cohort analysis rather than for time-series analysis. The second problem is that in the original VIIRS product (‘vcmcfg’) a lot of images are contaminated by stray-light (Mills et al., 2013). Stray-light is light that unexpectedly reaches the sensor during image formation due to design failure. It is indicative that there are no data for June over Europe for the original product due to stray-light. Data are missing for other months also, depending on the area. Therefore, the analysis of the summer period is impossible, given the quite extensive absence of data. For this reason, the alternative product (‘vcmslcfg’) is used here with radiance vales undergone a stray-light correction procedure (Mills et al., 2013). Consequently, missing values is less of a problem.

Greece is selected as the specific country for method validation. It receives major flows of tourists, particularly during the summer. The annual amount of tourists is approximately 23.5 million, roughly two times its resident population. At the same time, tourism accounts for almost one-fifth of the national GDP. In addition, ownership of second-houses is quite common. Approximately one third of citizens owns a second house. Hellenic National Statistical Authority statistics have been used as reference for the validation of the analysis. This is the source for the resident population in 2011 census (http://www.statistics.gr/el/statistics/-/publication/SAM03/2011) as well as arrivals by sea, aggregated in trimesters, (http://www.statistics.gr/el/statistics/-/publication/SMA06/2016-Q4). The monthly counts of arrivals to the islands by air, were obtained from the Hellenic Civil Aviation Authority (http://www.ypa.gr/en/profile/statistics/yearstatistics).

3. Method

3.1. Measuring seasonality by remote sensing

The first step of the method is to estimate the amount of average lights per month, for each region. The commonly used Sum of Lights (SoL) index is adopted (Elvidge et al., 2014):

$$SoL = \sum_{i} DN_i, \text{ for } DN > 2 \tag{1}$$

where DN_i are the digital number values of night-lights within a specific region.

Small values are considered to be background noise and excluded from calculation. SoL is used as a proxy of total (resident and seasonal) ambient population at a given time. The average SoL per month is then calculated for the three years that VIIRS night-lights are available in order to reduce annual fluctuations due to weather, temporal lights and other noise. Unlike OLS, VIIRS data has not undergone ephemeral lights removal processing.

$$SoL_{mean}^{m} = \frac{(SoL_{2014}^{m} + SoL_{2015}^{m} + SoL_{2016}^{m})}{3} \tag{2}$$

where m in [1, 12] being each month of the year.

The process is shown, for one island as an example, in Fig. 1. It is evident there that nighttime intensities for different months are somewhat different across the three years. The divergence is larger during the summer period, presumably because it is more affected by tourism. There are three main reasons for the fluctuations (i) national and international factors affecting tourism such as the financial crisis, the attempted coup in Turkey during July 2016, the different phases of the Syrian conflict and refugees since 2011 etc. (ii) holidays with moving dates, especially easter holidays (iii) noise in the data already mentioned such as stray-light, ephemeral lights etc. The proposed method tries to deal with the first two factors by year averaging and with the third by filtering. In this study area, VIIRS data are not expected to be significantly affected by weather conditions. On one hand cloud-cover is excluded by data design, on the other hand the amount of days with snow coverage in this part of the globe is negligible.

Then, a quantity that we will call observed 'seasonality coefficient' is calculated as

$$S_{obs}^{m} = \frac{SoL^{m}}{SoL^{March}} \tag{3}$$

again with m in [1, 12] being each month of the year.

The rational of dividing each month's value with that of March lies in that month being generally considered to exhibit the lowest seasonal activity. Therefore, March values should be regarded as the closest to the resident population alone. This is the reason why March is normally the census month in Greece. In effect s_{obs} shows how many times the activity of a specific month is higher compared to March. In other words, the activity of a given month is expressed as a function of the resident population, as perceived by night-lights. An example is shown in Table 2 for the selected island. In this example, the average activity in August is four times-March. The three annual time-series are reasonably consistent but there are instances of irrational variations for the same month, from one year to the next. Unreasonably low values are sometimes present due to missing values, because of stray-light contamination. Unreasonably high values are also sometimes present in the data, mainly due to random special events such as forest fires etc. Therefore, three rules where applied to filter out as much noise as possible:

rule 1: if $SoL_{obs}^{m} < \frac{SoL_{obs}^{median\ m1-m12}}{3}$ then $SoL_{obs}^{m} = null$, $SoL_{obs}{}^{median}{}_{m1-m12}$ is the median of all months and years

rule 2: if $SoL_{obs}{}^{m} > 2 * SoL_{obs}{}^{median\ m}$ then $SoL_{obs}{}^{m} = null$

rule 3: If $SoL_{March} = null$, then use $SoL_{February}$ or SoL_{April} in this order, in Eq. (3).

Obviously these rules are subjective. Nevertheless, it is evident in the results that the application of these rules effectively cleans the data from most of the superficial fluctuation. Other filtering methods are discussed latter on in the text.

Based on the seasonality coefficient (s_{obs}), three additional features are derived:

i) the peak month, i.e. the month when s_{obs} reaches its maximum value,
ii) the peak value, i.e. actual s_{obs} value of the peak month.
iii) the season's length, defined as the number of months where $s_{obs} > s_{obs}{}^{March} * 1.5$

The threshold value of 1.5 is chosen by visual examination of the data in order to avoid noise and especially the Christmas effect. This is a peak in December, in most spatial units, that can only be explained by illuminating decorations for Christmas celebrations. The effect is visible in Fig. 1 and Table 1. The actual threshold value is not critical for the analysis as the same number is applied to all spatial units and it is mainly used for visual inspection.

3.2. Validation – measuring seasonality by national statistics

The validation of results is quite difficult for the simple fact that national sources lack seasonal population statistics. This highlights at

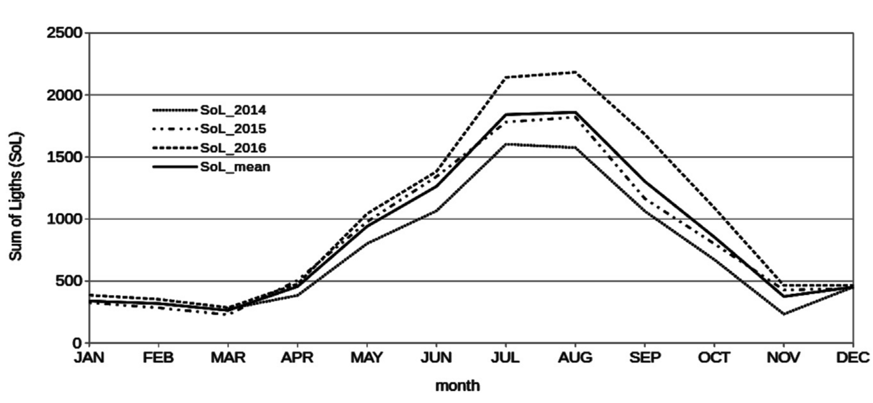

Fig. 1. Seasonal variation of SoL for one example – Myconos island.

Table 2
Processed SoL values for Myconos island as an example.

.	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
2014	701	547	593	770	1300	1627	2154	2138	1583	1117	614	890
2015	630	618	506	884	1436	1795	2312	2383	1666	1262	868	837
2016	693	653	583	879	1559	1901	2649	2764	2239	1662	975	1092
SoL^{mean}	675	606	561	844	1432	1774	2372	2428	1829	1347	819	940
s_{obs}	1.2	1.1	1.0	1.5	2.6	3.2	4.2	4.3	3.3	2.4	1.5	1.7

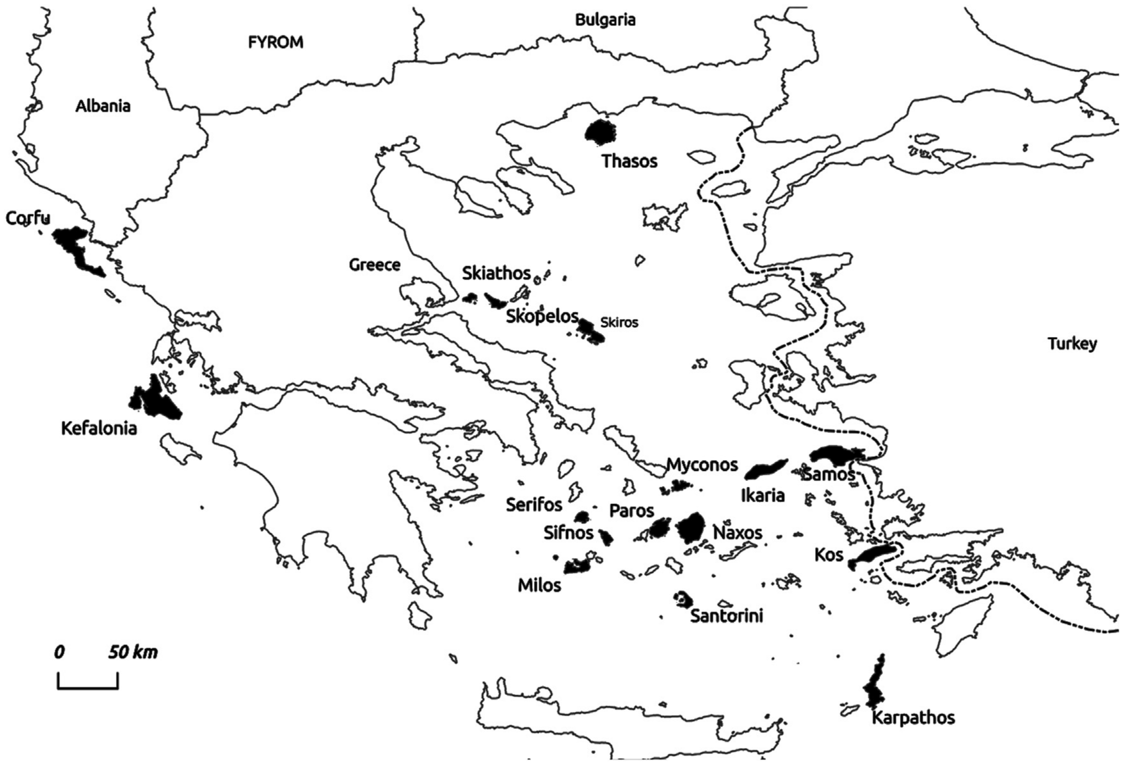

Fig. 2. Validation islands.

the same time the importance of establishing a method to yield seasonality estimates. The procedure to validate the results is to select a number of islands as reference, shown in Fig. 2, where it is much easier to capture the flows of people. The basic characteristics of the validation islands are summarized in Table 3.

Fortunately, monthly arrivals by air and by sea are recorded by the national authorities. Arrivals by sea, recorded in trimesters, have been disaggregated to monthly estimates by matching the distribution of arrivals by air. The sum of air and sea arrivals yields the total arrivals per month shown in Table 4. Note that some islands receive tourist counts corresponding to many times their resident population (e.g. 8.5 times for Kos in August). Arrivals have been averaged per month for the same three years as with the night-lights. Overall, arrivals by boat and air should be more reliable compared to occupancy of touristic accommodation data because arrivals are linked to security and safety controls, thoroughly performed by the police and the coastguard, leading to negligible if any ambiguity in recording the data.

In a similar manner, as we did with the night-lights, seasonality coefficient is estimated, this time based on official statistics. It is assumed that the total population on each island, at any given day, will be

$$P_t = P_s + P_r \tag{4}$$

with P_t the total daily population, P_s the average daily seasonal population and P_r the resident population (2011 census).

Table 3
Basic characteristics of validation islands.

Islands	Size (km^2)	Resident population 2011	Count of VIIRS pixels
Icaria	321.97	8,423	1,495
Sifnos	98.29	2,625	453
Thasos	508.69	13,770	2,373
Kefalonia	1,001.13	35,801	4,647
Kos	360.19	33,388	1,673
Skopelos	123.96	4,960	583
Myconos	133.54	10,134	617
Corfu	767.21	102,071	3,571
Naxos	542.33	17,930	2,511
Santorini	112.69	15,550	524
Paros	249.24	13,715	1,166
Milos	209.66	4,977	972
Karpathos	398.72	6,226	1,857
Skiathos	63.91	6,088	300
Serifos	95.82	1,420	442
Samos	607.59	32,977	2,822
Skyros	283.73	2,994	1,330
Min	63.913	1,420	300
Max	1,001.13	102,071	4,647
Mean	345.80	18,414.65	1,608
σ	265.50	24,232.56	1,233.18

Table 4
Average arrival values (air and sea).

Islands	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Icaria	3,087	2,473	3,214	6,028	6,466	8,567	19,602	21,264	12,599	4,854	3,245	3,572
Sifnos	1,059	1,217	1,744	7,010	8,134	7,980	23,792	23,579	18,689	2,799	1,211	1,061
Thasos	17,100	18,775	25,593	3,784	78,200	140,197	177,589	200,563	108,980	54,510	12,758	10,754
Kefalonia	32,307	37,123	43,459	14,349	102,374	164,972	217,150	219,570	146,034	88,153	17,856	16,236
Kos	17,441	16,067	21,154	33,122	155,510	215,012	283,231	284,027	202,749	106,238	8,636	8,744
Skopelos	1,991	2,167	2,950	469	9,694	17,382	31,153	35,188	19,119	5,880	1,359	1,150
Myconos	10,471	11,496	17,624	27,494	97,897	177,682	225,990	240,199	122,899	46,678	8,257	6,709
Corfu	40,228	39,178	53,334	59,075	217,270	332,549	391,409	386,689	272,633	169,248	21,377	20,346
Naxos	9,382	9,072	12,413	20,551	32,722	56,294	81,750	82,435	64,822	17,905	8,123	7,273
Santorini	17,043	19,596	32,068	70,049	161,618	224,467	281,222	287,891	206,773	78,291	22,830	16,212
Paros	11,635	13,368	19,264	51,962	62,016	62,196	162,585	161,176	127,738	27,935	12,094	10,589
Milos	2,326	2,215	2,653	6,306	10,863	17,363	36,650	36,262	28,540	4,572	2,873	2,469
Karpathos	2,009	1,596	2,136	3,001	9,598	19,869	28,450	29,569	19,007	3,789	1,833	1,833
Skiathos	3,342	3,655	4,970	1,576	32,554	58,362	83,644	94,465	51,329	9,538	2,220	1,874
Serifos	649	745	1,071	3,981	4,643	4,602	14,187	14,063	11,146	1,752	756	664
Samos	8,096	15,043	8,015	10,418	27,101	40,904	54,471	55,903	40,701	13,894	8,701	7,602
Skyros	2,248	3,449	3,354	3,945	5,291	9,634	17,113	18,065	8,074	3,447	2,428	2,091

NB for JUL–DEC only data for 2014–5 were available at the time of analysis.

An estimate of the seasonal population based on available data can be

$$P_s = \frac{a*d}{30} \tag{5}$$

with P_s the daily seasonal population and a the total arrivals per month (air and sea). The denominator is the average days a month has, in order to convert to daily estimates. Parameter d is the 'average length of stay of visitors', i.e. the days spent by each visitor on the island. Based on the UNCTAD (2007) Handbook of Statistics $d = 5.37$ Because of second-houses and because the contribution of second-houses on average length of stay varies per region this d value is meaningful only as a starting point. Therefore values $d \in [2, 10]$ were tested. Overall, validation results, in terms of Pearson's correlation, show little sensitivity in the setting of d. However, the estimation of absolute seasonal population is heavily affected (as evident later in Eq. (7)). The seasonal coefficient this time can be derived as:

$$s_{\text{ref}} = \frac{P_s + P_r}{P_r} = \frac{\frac{a*d}{30} + P_r}{P_r} \tag{6}$$

with s_{ref} the seasonality coefficient based on official statistics, a the total arrivals per month, d the average length of stay of visitors and P_r the resident population.

The seasonality coefficient s_{ref} based on the census data for some of the islands is shown on Fig. 3. Different seasonal patterns are evident in terms of how spread the season is as well as how rapid the season goes from one level to the other.

The overall objective of the seasonality coefficient s is to express the total population of a spatial unit in a given month as a function of its resident population (March), when it is assumed that there is no seasonal population $s = 1$ and $P_t = P_r$. Absolute estimates of total monthly population are obtained via the following equation:

$$P_t = s_{\text{obs}}*P_r \tag{7}$$

The seasonality coefficient of Table 4 is shown on Table 5.

3.3. Validation – comparison of remote sensing and statistics estimates

We have obtained two independent seasonality coefficient estimates. One by night-lights (s_{obs}) and a reference one by official data (s_{ref}). The Pearson's correlation of the two estimates is calculated to evaluate the night-lights method. The average Pearson's correlation for all the islands combined, with an average days of stay $d = 5.37$, is 0.85. For $d \in [2, 10]$ the Pearson's correlation varies ± 0.02 of this value. Individual correlation values are shown in Table 5. A diagram of the two independent estimates for the test island is presented in Fig. 4.

As an example of deriving absolute total population estimates, Myconos in August has $s = 4.33$ (Table 5.a) and knowing from 2011 census that its resident population was 10,134 inhabitants, Eq. (7) yields $P_T = S * P_r = 4.33 * 10{,}134 = 43{,}880$ people.

4. Results

The method that was validated in the selected islands was applied to all municipalities of the country. The three derived seasonality quantities are mapped on Figs. 5, 6, and 7. Fig. 5 shows the peak month,

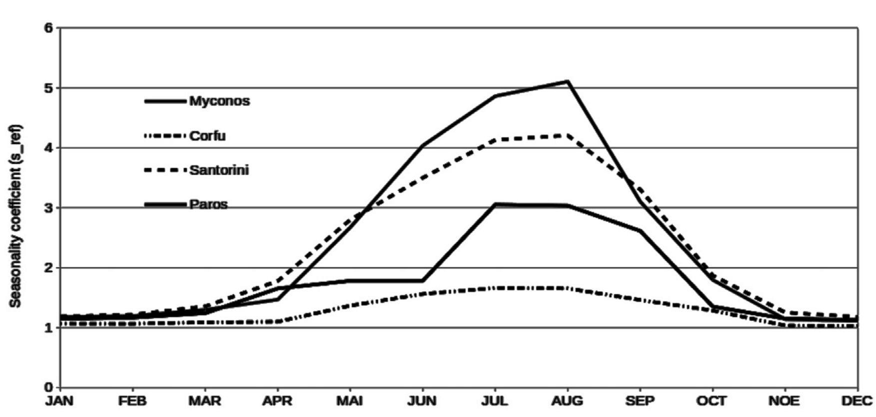

Fig. 3. Seasonality coefficient s_{ref} based on national statistics for some islands ($d = 5.37$).

Table 5
Seasonal coefficients s_{obs} and s_{ref} and their Pearson's correlation.

(a) s_{obs} derived by night-lights

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Icaria	1.23	0.77	1.00	1.09	1.18	1.27	1.79	2.20	1.18	1.06	0.92	1.26
Sifnos	1.47	1.25	1.00	1.47	2.26	2.79	3.64	4.78	2.64	1.58	1.72	2.24
Thasos	1.13	1.02	1.00	1.05	1.18	1.20	1.20	1.27	1.23	1.11	1.08	1.27
Kefalonia	1.28	1.17	1.00	1.19	1.43	1.54	1.71	1.88	1.48	1.18	1.17	1.66
Kos	1.06	1.11	1.00	1.16	1.71	1.80	1.83	1.66	1.62	1.55	1.20	1.26
Skopelos	2.01	1.25	1.00	1.04	1.68	2.10	2.77	3.24	2.19	1.38	1.53	2.44
Myconos	1.20	1.08	1.00	1.51	2.55	3.16	4.23	4.33	3.26	2.40	1.46	1.68
Corfu	1.10	0.94	1.00	1.07	1.30	1.44	1.46	1.50	1.31	1.13	1.06	1.37
Naxos	1.22	1.03	1.00	1.11	1.23	1.55	1.73	1.86	1.37	1.21	1.09	1.39
Santorini	1.04	1.06	1.00	1.37	1.71	2.09	2.30	2.27	2.00	1.82	1.31	1.32
Paros	1.18	0.87	1.00	1.26	1.49	1.85	2.75	3.08	1.81	1.33	1.06	1.47
Milos	1.06	0.98	1.00	1.60	1.72	1.83	2.06	2.16	1.61	1.34	1.14	1.26
Karpathos	1.35	1.10	1.00	1.30	1.08	2.28	2.54	2.54	2.14	1.61	1.22	1.67
Skiathos	1.19	0.97	1.00	1.04	1.45	1.58	1.81	1.84	1.33	1.07	1.01	1.08
Serifos	1.03	0.73	1.00	1.14	1.90	2.34	3.01	3.81	1.74	1.25	1.06	1.14
Samos	1.07	1.02	1.00	1.01	1.12	1.22	1.49	1.16	1.12	1.15	1.12	1.20
Skyros	1.76	1.35	1.00	1.37	1.99	2.56	3.12	5.24	2.35	1.63	2.06	2.15

(b) s_{ref} by national statistics (d = 5.37)

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Pearson's cor. ($s_{obs} - s_{ref}$)
Icaria	1.06	1.05	1.07	1.12	1.13	1.18	1.40	1.44	1.26	1.10	1.07	1.07	0.90
Sifnos	1.07	1.08	1.11	1.46	1.54	1.53	2.57	2.56	2.23	1.18	1.08	1.07	0.87
Thasos	1.21	1.24	1.32	1.05	1.98	2.76	3.23	3.52	2.37	1.69	1.16	1.13	0.65
Kefalonia	1.16	1.18	1.21	1.07	1.49	1.80	2.05	2.06	1.71	1.43	1.09	1.08	0.73
Kos	1.09	1.08	1.11	1.17	1.81	2.12	2.47	2.47	2.05	1.55	1.04	1.04	0.91
Skopelos	1.07	1.08	1.10	1.02	1.34	1.61	2.09	2.23	1.67	1.20	1.05	1.04	0.80
Myconos	1.18	1.19	1.30	1.47	2.67	4.04	4.86	5.11	3.10	1.80	1.14	1.11	0.96
Corfu	1.07	1.07	1.09	1.10	1.37	1.56	1.66	1.66	1.46	1.29	1.04	1.03	0.82
Naxos	1.09	1.09	1.12	1.20	1.32	1.54	1.79	1.80	1.63	1.17	1.08	1.07	0.87
Santorini	1.19	1.22	1.36	1.78	2.80	3.50	4.13	4.21	3.30	1.87	1.25	1.18	0.94
Paros	1.15	1.17	1.24	1.66	1.78	1.79	3.05	3.04	2.61	1.35	1.15	1.13	0.92
Milos	1.08	1.08	1.09	1.22	1.38	1.60	2.27	2.26	1.99	1.16	1.10	1.09	0.87
Karpathos	1.06	1.04	1.06	1.08	1.26	1.55	1.79	1.82	1.53	1.10	1.05	1.05	0.90
Skiathos	1.09	1.10	1.14	1.04	1.93	2.66	3.38	3.69	2.46	1.27	1.06	1.05	0.96
Serifos	1.08	1.09	1.13	1.49	1.57	1.56	2.73	2.72	2.36	1.21	1.09	1.08	0.88
Samos	1.04	1.08	1.04	1.05	1.14	1.21	1.29	1.29	1.21	1.07	1.05	1.04	0.65
Skyros	1.13	1.20	1.19	1.23	1.31	1.56	1.99	2.04	1.47	1.20	1.14	1.12	0.86

when seasonality gets its maximum value. Fig. 6 shows the maximum value, observed during the peak month. Fig. 7 shows the length of the season defined as how many months the seasonality coefficient observed is 50% higher than that of March. The maps are in agreement with the expected trend that coastal areas tend to peak around summer whereas mainland areas during winter and spring. In less touristic areas, with little annual variation, the Christmas effect has stronger relative impact. Peak values do not necessarily reveal the most touristic places but rather the areas that receive the most seasonal population compared to their resident population. Season length highlights both

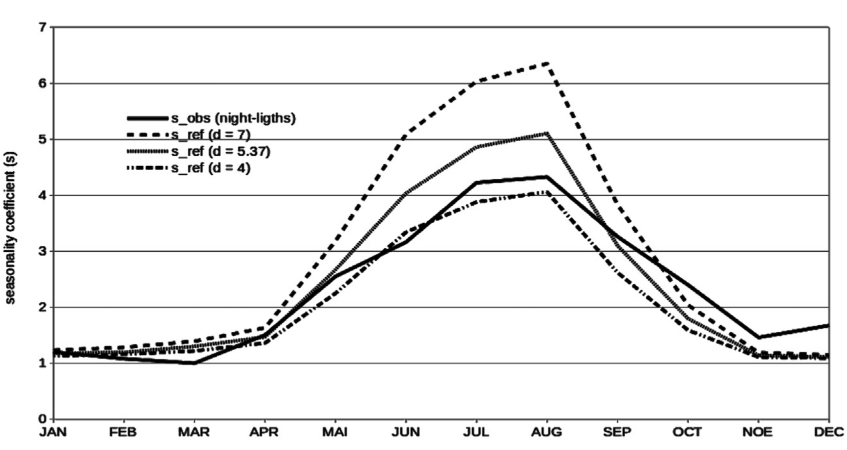

Fig. 4. Comparing the two independent estimates of the seasonality coefficient for Myconos.

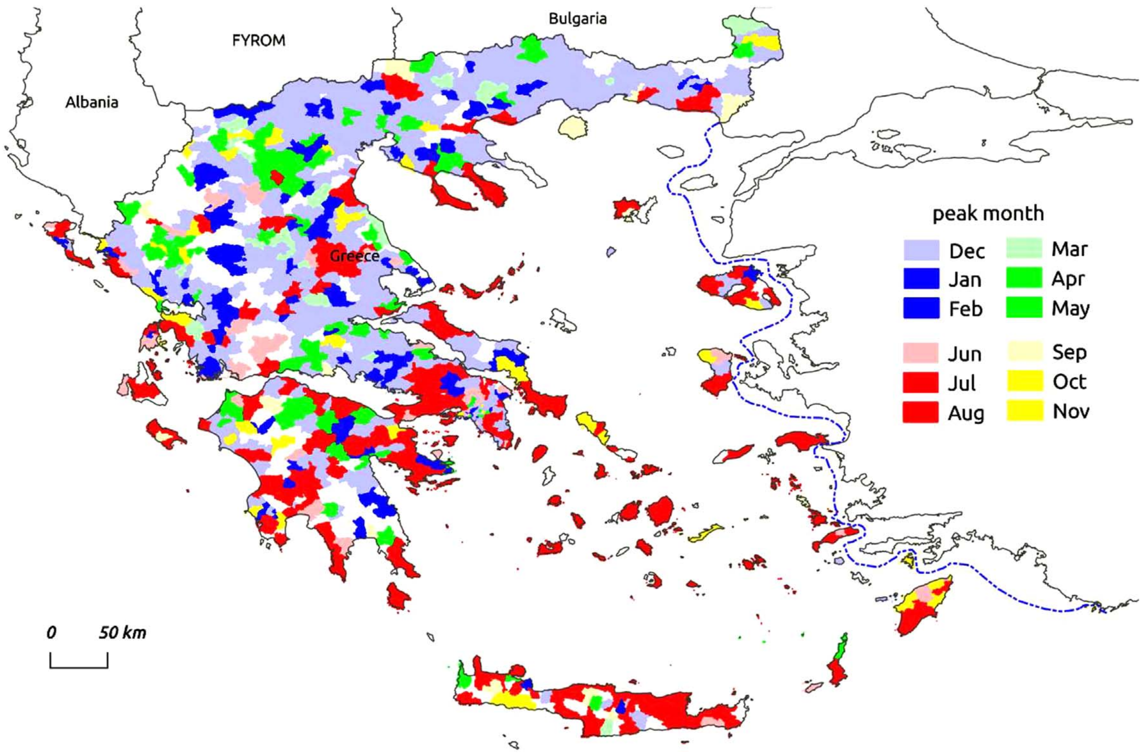

Fig. 5. Peak month.

coastal touristic areas and winter destinations on the map.

The spatial level of analysis is very important. The choice of these relatively fine areal units, reveals the actual magnitude of the phenomenon, which is very unevenly distributed in space. Due to the modifiable areal unit problem (MAUP) the level of aggregation substantially affects results. In specific, the effect of seasonality disappears when using large administrative units, that tend to average it out.

5. Discussion

The main advantage of the remote sensing approach, compared to the cell-phone data alternative, is that night-lights data are global and public domain. The proposed method is based at the most detailed open domain night-lights data available.

Things to consider regarding the method include first of all the noise

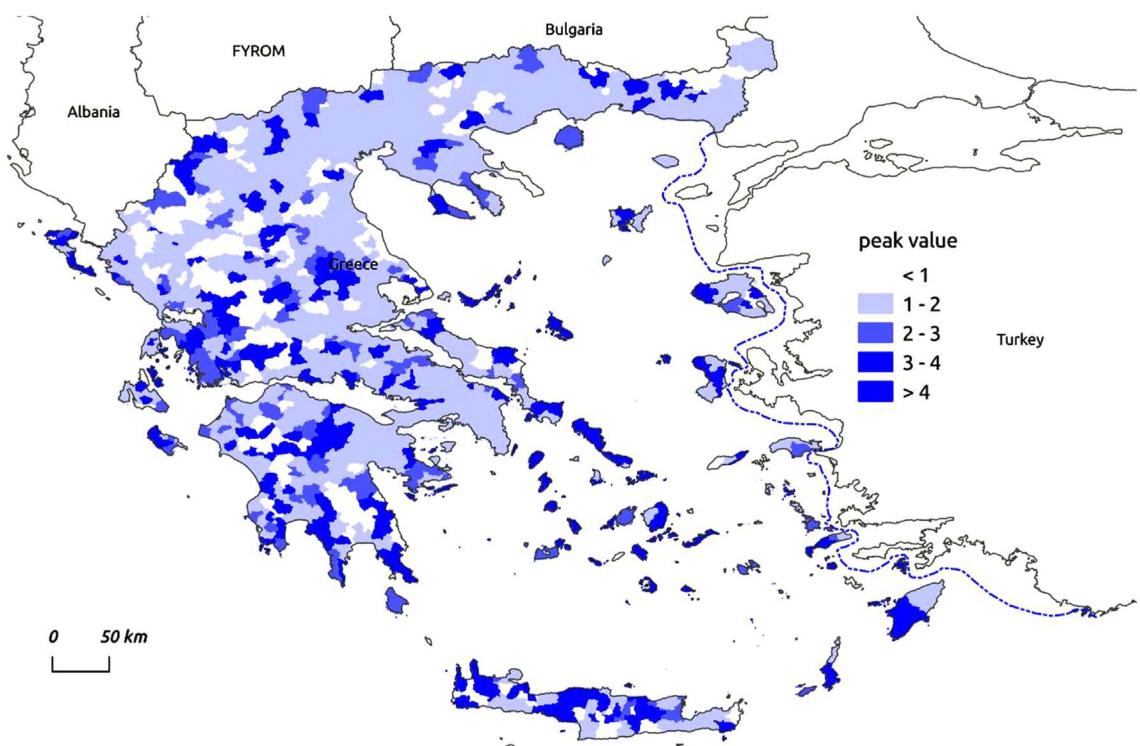

Fig. 6. Peak value.

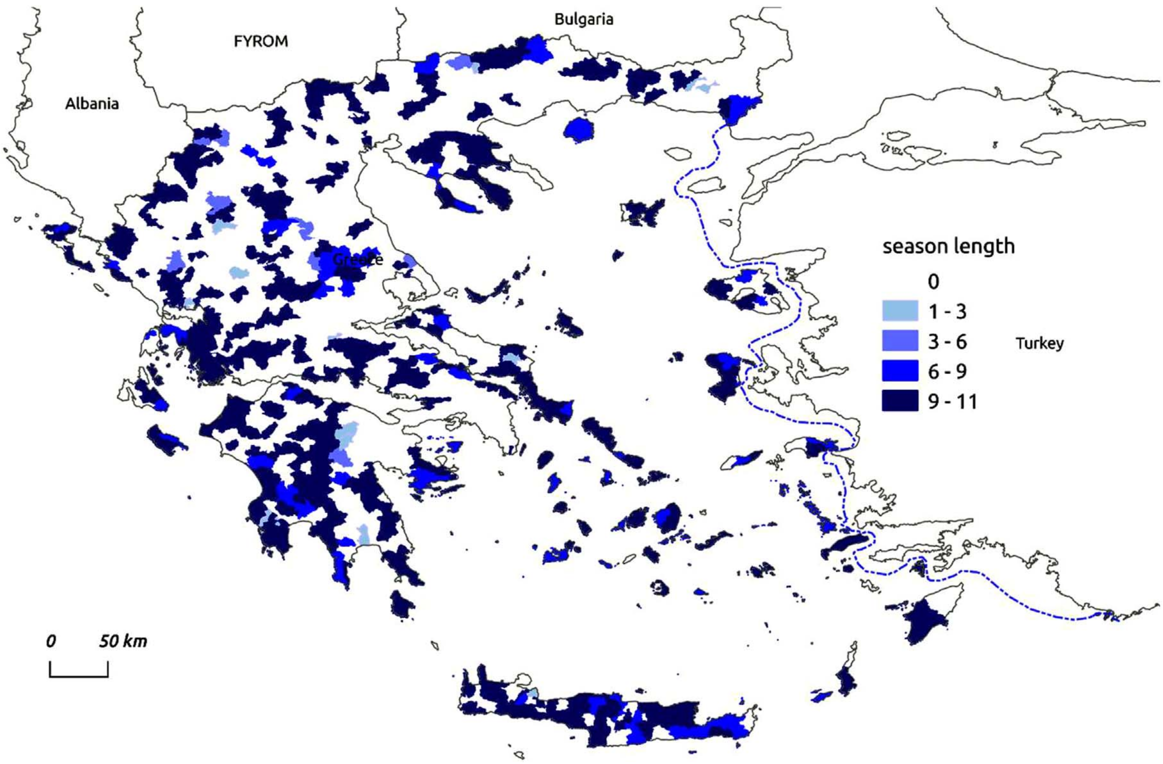

Fig. 7. Season length.

present in the VIIRS night-lights. The stray-light corrected version is consistent enough to be used for seasonal population estimates. Ephemeral lights are removed here using rules but other methods are also available in the literature. The most prominent of which is using OLS as a mask, given that OLS has ephemeral lights removed (Li, Xu, Chen, & Li, 2013). This is an open field for future research.

Also, some thresholds in the process have been determined by trial and error, such as the one used to identify the peak month. As explained earlier, these thresholds are justified because the same value is applied to all spatial units concurrently. Scaling factors have also been used. The most important of which is the mean length of stay. A starting value has been obtained from the literature and a range of values has then been explored to test for sensitivity. Overall the setting of this scaling factor was found to be stable in terms of correlation of observed with reference values.

In addition, using one value as the mean length of stay for all geographical areas and all times within the year is a generalization that introduces some error. The mean length of stay is most likely different per category of visitor (e.g. hotel guests, second home owners, seasonal workers) as well as within the year and across space. However, the strong correlation of the observed and the reference values for the test islands indicate that this generalization does not have the power to significantly affect the accuracy of estimating the population stock actually present on an island at any time. The stability of the results probably has to do with the fact that tourists consist of a much larger proportion of the seasonal population compared to all other groups (seasonal workers, second-house owners).

Moreover, 2011 census resident population is used as a constant for all the years in this study (2014–2016), in the absence of more recent data available. Although there is some mismatch due to this fact, the change of residence population within a few years is typically negligible. At national scale this change is as low as 3%.

The validation performed is somehow limited because it was done by selecting a relatively small number of islands as reference. There is no easy way to validate the method at the national scale because there are no suitable data at that scale and also because of the averaging-out effect already mentioned. Future work is required to assess the validity of the approach in non-island areas provided that suitable reference data exist.

The method would probably not perform well in areas with extensive snow cover in the winter as year-to-year fluctuations could be attributed to weather rather than actual population changes. Also in areas with very low seasonal variations the changes recorded by night-lights are more likely to be attributed to celebration decorations rather than seasonal population. These low activity areas are also more prone to be affected by ephemeral events (forest fires, black-outs etc.) that are not removed a priori in VIIRS.

It is understood that the emission of night-lights is not only a factor of population size but also of population welfare levels, the state of infrastructure etc. In Greece, the poorest region's GDP amounts to approximately half that of the richest. While the setting is definitely not everywhere identical, the proposed analysis is valid because it is based on comparing the same region at different times rather than comparing different regions. Comparison across regions comes only when the seasonality coefficient has been estimated. The seasonality coefficient removes regional differences because each area is scaled by its own lowest activity.

Furthermore, arrival data could include commuting travels of local residents to the mainland and vice versa but in practice these numbers are negligible. The reason is that travel times from the islands to the mainland are so long that commuting is close to impossible. For example, for the cluster of islands in the Aegean (Fig. 2, south-east quadrant), one-way travel time to Athens by ferry boat ranges from three to seventeen hours. Weather conditions, limited service during the winter and ticket costs render the choice of systematic commuting improbable. The cost of travel by air is prohibitive for the vast majority of people and not all islands have and airfield. Travel between islands for business and shopping is possible but limited to the very necessary due to the associated time and cost. Illegal arrivals are not recorded in the reference data but their amount is expected to be negligible for the islands selected for validation. Arrivals by private boats are also not recorded but it is expected that their impact is proportionally small given that the capacity of the marina's is relatively low.

Finally, in the proposed model visiting and resident population are treated as equal contributors to nighttime light emissions. This is also a generalization given that light emissions attributed to different population groups are likely to be different. Installed public street lights operate in a stable manner throughout the year. The additional light associated with visiting population comes from private houses and streets that are closed when the season is off, hotels and businesses that are closed during the winter, and the increased car lights captured by the sensor.

6. Conclusion

Overall, the newly available VIIRS night-lights provide an unprecedented opportunity to map monthly variations of night-light and therefore derive estimates of seasonal ambient population. The core method could be transferred to other countries as this approach can be used for both summer and winter population variations. Nevertheless, different scaling factors should presumably be used based on local religious celebrations, weather conditions (extensive snow cover) etc. Such estimates are very useful in a wide range of scientific fields including planning. In case more frequent night-time data become available in the future, e.g. weekly or daily, the method could be used to derive public domain seasonal population estimates for a totally new range of applications such as commuting patterns and migration due to conflicts.

Acknowledgements

VIIRS composites are a product generated by the Earth Observation Group, NOAA National Geophysical Data Center. The first author was supported for this work as a Fulbright Fellow – Greece (visiting scholar 2016-7). The second author was supported by the Hellenic State Scholarships Foundation (IKY-SIEMENS contract no. 2016-017-0173-10841).

References

Airbnb. About us. (2017). URL https://www.airbnb.com/about/about-us (accessed on 8 August 2017).

Amaral, S., Camara, G., Miguel Vieira Monteiro, A., Alberto Quintanilha, J., & Elvidge, C. (2005). Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. *Computers, Environment and Urban Systems, 29*, 179–195.

Bharti, N., Tatem, A., Ferrari, M., Grais, R., Djibo, A., & Grenfell, B. (2011). Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science, 334*, 1424–1427.

Coccossis, H., & Mexa, A. (2004). *The challenge of tourism carrying capacity assessment theory and practice.* 312 (Routledge).

Coccossis, H., & Parpairis, A. (2000). Tourism and the environment: Some observations of the concept of carrying capacity. *Tourism and the environment. 6. Tourism and the environment* (pp. 91–105). Springer Netherlands.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F., Gaughan, A., ... Tatem, A. (2014). Dynamic population mapping using mobile phone data. *PNAS, 111*(45), 15888–15893.

Elvidge, C., Hsu, F.-C., Baugh, K., & Gosh, T. (2014). National trends in satellite-observed lighting 1992–2012. In Q. Weng (Ed.). *Global urban monitoring and assessment through earth observation* (pp. 97–120). CRC Press.

Erbach-Schoenberg, E., Alegana, V., Sorichetta, A., Lianard, C., et al. (2016). Dynamic denominators: The impact of seasonally varying population numbers on disease incidence estimates. *Population Health Metrics, 14*, 35. http://dx.doi.org/10.1186/s12963-016-0106-0.

Hanaoka, K. (2016). New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mobile phone users' locations. *Environment and Planning. B, Planning & Design,* 17. http://dx.doi.org/10.1177/0265813516672454.

Kloog, I., Haim, A., Stevens, R., & Portnov, B. (2009). Global co-distribution of light at night (LAN) and cancers of prostate, colon and lung in men. *Chronobiology International, 26*(1), 108–125. http://dx.doi.org/10.1080/07420520802694020.

Li, X., & Li, D. (2014). Can night-time light images play a role in evaluating the Syrian Crisis? *International Journal of Remote Sensing, 35*(18), 6648–6661. http://dx.doi.org/10.1080/01431161.2014.971469.

Li, X., Xu, H., Chen, X., & Li, C. (2013). Potential of NPP-VIIRS nighttime light imagery for modeling the regional economy of China. *Remote Sensing, 5*, 3057–3081.

Ma, T., Zhou, C., Pei, T., Haynie, S., & Fan, J. (2012). Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities. *Remote Sensing of Environment, 124*, 99–107.

Mills, S., Weiss, S., & Liang, C. (2013). VIIRS day/night band (DNB) stray light characterization and correction. In: *Proc. SPIE 8866, Earth Observing Systems XVIII, 88661P* (pp. 18). . http://dx.doi.org/10.1117/12.2023107.

Ratti, C., Frenchman, D., Pulselli, R., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning. B, Planning & Design, 33*, 727–748.

Reades, J., Calabrese, F., & Ratti, C. (2009). Eigenplaces: Analysing cities using the space and time structure of the mobile phone network. *Environment and Planning. B, Planning & Design, 36*, 824–836.

Roman, M. O., & Stokes, E. C. (2015). Holidays in lights: Tracking cultural patterns in demand for energy services. *Earth's Future, 3*, 182–205.

Smith, A., Newing, A., Quinn, N., Martin, D., Cockings, S., & Neal, J. (2015). Assessing the impact of seasonal population fluctuation on regional flood risk management. *ISPRS International Journal of Geo-Information, 4*, 1118–1141.

Stathakis, D. (2015, July). *Deriving soil sealing degree from nighttime VIIRS imagery*. Milan, Italy: IGARSS4. http://dx.doi.org/10.13140/RG.2.1.4886.4403.

Stathakis, D. (2016). Intercalibration of DMSP/OLS by parallel regressions. *IEEE Geoscience and Remote Sensing Letters, 13*(10), 1420–1424.

Stathakis, D., Tselios, V., & Faraslis, I. (2015). Urbanization in European regions based on night lights. *Remote Sensing Applications: Society and Environment, 2*, 26–34.

Sutton, P., Elvidge, C., & Obremski, T. (2003). Building and evaluating models to estimate ambient population density. *Photogrammetric Engineering and Remote Sensing, 69*, 545–554.

Triantakonstantis, D., & Stathakis, D. (2014). Cokriging areal interpolation for estimating economic activity using night-time light satellite data. In B. Murgante, (Ed.). *Computational science and its applications – ICCSA* (pp. 243–252). Springer. http://dx.doi.org/10.1007/978-3-319-09147-1_18.

United Nations Conference on Trade and Development (UNCTAD). available on-line http://www.nationsencyclopedia.com/WorldStats/UNCTAD-average-length-stay-visitors.html (last access 10 May 2017).

Wesolowski, A., Metcalf, C., Eagle, N., Kombich, J., Grenfell, B., Bjornstad, Q., ... Buckee, C. (2015). Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *PNAS, 112*(35), 11114–11119.

Wilson, R., Erbach-Schoenberg, E., Albert, M., Power, D., et al. (2016, Feb 24). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLOS Currents Disasters*. http://dx.doi.org/10.1371/currents.dis.d073fbece328e4c39087bc086d694b5c.

Yang, X., Fang, Z., & Xu, Y. (2016). Understanding spatiotemporal patterns of human convergence and divergence using mobile phone location data. *International Journal of Geo-Information, 5*(10), 18.

Zervas, G., Proserpio, D., & Byers, J. (2016). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Journal of Marketing*. http://dx.doi.org/10.1509/jmr.15.0204 (in press).

Zhang, Q., & Seto, K. (2011). Mapping urbanization dynamics at regional and global scales using multi-temporal DMSP/OLS nighttime light data. *Remote Sensing of Environment, 115*, 2320–2329.

Zhang, Q., & Seto, K. (2013). Can night-time light data identify typologies of urbanization? A global assessment of successes and failures. *Remote Sensing, 5*(7), 3476–3494.