Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

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

A R T I C L E I N F O

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

A B S T R A C T S

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

Please cite this article as: Stathakis, D., Computers, Environment and Urban Systems (2017), https://doi.org/10.1016/j.compenvurbsys.2017.12.001

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 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.

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

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