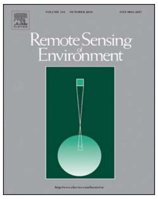
ARTICLE IN PRESS

Remote Sensing of Environment xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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



Census-independent population mapping in northern Nigeria☆

Eric M. Weber[a,*], Vincent Y. Seaman[b], Robert N. Stewart[a], Tomas J. Bird[c,d], Andrew J. Tatem[c,d], Jacob J. McKee[a], Budhendra L. Bhaduri[a], Jessica J. Moehl[a], Andrew E. Reith[a]

[a] *Urban Dynamics Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA*
[b] *Bill & Melinda Gates Foundation, Seattle, WA, USA*
[c] *WorldPop, Department of Geography and Environment, University of Southampton, Highfield, Southampton, UK*
[d] *Flowminder Foundation, Stockholm, Sweden*

ARTICLE INFO

Keywords:
Population
Settlement mapping
Nigeria
Demographics
Polio

ABSTRACT

Although remote sensing has long been used to aid in the estimation of population, it has usually been in the context of spatial disaggregation of national census data, with the census counts serving both as observational data for specifying models and as constraints on model outputs. Here we present a framework for estimating populations from the bottom up, entirely independently of national census data, a critical need in areas without recent and reliable census data. To make observations of population density, we replace national census data with a microcensus, in which we enumerate population for a sample of small areas within the states of Kano and Kaduna in northern Nigeria. Using supervised texture-based classifiers with very high resolution satellite imagery, we produce a binary map of human settlement at 8-meter resolution across the two states and then a more refined classification consisting of 7 residential types and 1 non-residential type. Using the residential types and a model linking them to the population density observations, we produce population estimates across the two states in a gridded raster format, at approximately 90-meter resolution. We also demonstrate a simulation framework for capturing uncertainty and presenting estimates as prediction intervals for any region of interest of any size and composition within the study region. Used in concert with previously published demographic estimates, our population estimates allowed for predictions of the population under 5 in ten administrative wards that fit strongly with reference data collected during polio vaccination campaigns.

1. Introduction

Current and spatially precise population estimates are a critical data input for efforts in governance, planning, and public health. Without an accurate count or estimate of the population denominator for an area, rates describing demographic compositions, births and deaths, disease incidence, health intervention coverage, technology penetration, service accessibility and voting turnout, for instance, are both difficult to measure and of limited value in future planning. More than one-third of the indicators established to measure progress on the United Nations (UN) Sustainable Development Goals (SDGs) (United Nations, 2016) are defined in terms of total population or a specific demographic subpopulation, despite the fact that the capacity to measure these denominators varies greatly from country to country, especially when data are needed for small areas, rather than at national or provincial levels.

One example of the critical need to ascertain populations for small areas can be found in the work of the Global Polio Eradication Initiative (GPEI) in Nigeria, which conducts regular vaccination campaigns with the aim of vaccinating every child under the age of five. Despite a host of innovative interventions in recent years (Vaz et al., 2016), the polio eradication effort in Nigeria has been hampered by areas of insecurity and a lack of access to all communities and children. The limited access, along with the inadequacy of available geodemographic data, make the accurate assessment of vaccination coverage a challenge, compromising the GPEI's ability to assess the effectiveness and efficacy of the vaccination campaigns (Barau et al., 2014). Even in the ideal case, when supplies, logistics, and freedom to operate allow access to all children in

---

☆ Copyright notice: This manuscript has been co-authored by one or more employees of UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

* Corresponding author.
E-mail addresses: weberem@ornl.gov (E.M. Weber), Vincent.Seaman@gatesfoundation.org (V.Y. Seaman), stewartrn@ornl.gov (R.N. Stewart), T.J.Bird@soton.ac.uk (T.J. Bird), A.J.Tatem@soton.ac.uk (A.J. Tatem), mckeejj@ornl.gov (J.J. McKee), bhaduribl@ornl.gov (B.L. Bhaduri), moehljj@ornl.gov (J.J. Moehl), reithae@ornl.gov (A.E. Reith).

http://dx.doi.org/10.1016/j.rse.2017.09.024
Received 20 February 2017; Received in revised form 6 September 2017; Accepted 16 September 2017

Please cite this article as: Weber, E.M., Remote Sensing of Environment (2017), http://dx.doi.org/10.1016/j.rse.2017.09.024

all areas, not knowing where all vaccine-eligible children reside can lead to children being missed by the campaigns. Similarly, the effectiveness of the routine immunization services provided by local health districts cannot be measured without an accurate target population denominator.

For locating and quantifying the number of vaccine-eligible children, national census data have limitations. The last national census in Nigeria occurred in 2006 and provided counts of the total population as well as the populations by sex and 5-year age groups at the level of the Local Government Area (LGA). This level of aggregation did not allow determination of the population of individual settlements within LGAs, a problem the National Population Commission acknowledged and attributed to the lack of authoritative lists and maps of localities (National Population Commission, 2009). Now, a decade removed from the census, ascertaining the population of large or small areas in Nigeria is even more problematic, as the differential growth rates among LGAs over that time is not accounted for in tabular projections using constant growth rates. The case of Nigeria is far from unique, and it is representative of the challenges faced by governments and NGOs attempting to implement ambitious programs in countries where the availability of detailed geographic and demographic data is inadequate (Tatem and Linard, 2011).

Although settings without recent and reliable census data are common, most research in spatially precise population estimation relies on national census data for observations of population counts. A common approach is to estimate a population density for each class of land cover or land use, whether by regressing the census populations on the areas of the different land classes (Fisher and Langford, 1995; Goodchild et al., 1993; Langford et al., 1991; Yuan et al., 1997) or by compiling an empirical sample for each class by identifying enumeration units that are completely (or mostly) covered by a single class (Mennis, 2003; Mennis and Hultgren, 2006). The census data can also be used to constrain estimates so that sums are preserved within the enumeration units. Whether this constraint is imposed depends on whether the goal of the estimation is a real interpolation of the census counts or predictions outside of the context of model training, whether for different regions or dates (Wu et al., 2005). Further refinements of census-based methods include incorporation of additional ancillary data in combination with land cover (Dobson et al., 2000; Stevens et al., 2015) and the application of alternative spatial denominators (other than area), such as building volume (Sridharan and Qiu, 2013), street lengths (Reibel and Bufalino, 2005), or residential address points (Zandbergen, 2011).

While the land classifications used in some early population estimation work were hand-drawn and guided by “controlled guesswork” (Wright, 1936), most modern techniques use data derived via remote sensing. Although a variety of remote sensing data and methods have been applied to population estimation problems, the increasing availability of high-resolution optical and radar imagery has contributed to a gradual trend, recognized at least as early as 2004 (Tatem and Hay, 2004), toward window-based textural classifications, which have been shown to be well suited for identifying and characterizing the complex structures of human settlements (Cheriyadat et al., 2010; Martino et al., 2003; Pesaresi, 2000). (Unless otherwise noted, our discussion of resolution throughout the text refers to spatial resolution.) In order to deploy these principles at regional and global scales, scalable workflows have been developed within computational platforms such as the Global Human Settlement (GHS) framework at the Joint Research Centre (JRC) of the European Commission (Pesaresi et al., 2013), the Settlement Mapper Tool (SMT) platform developed at Oak Ridge National Laboratory (ORNL) (Cheriyadat et al., 2007; Patlolla et al., 2012), and the German Aerospace Center (DLR)'s Urban Footprint Processor (UFP) (Esch et al., 2013). The highest-resolution settlement layer with global coverage from these platforms is currently DLR's Global Urban Footprint (GUF), which can be licensed at 12-meter resolution for scientific use. Higher resolutions of 10 m (Florczyk et al., 2016) and 8 m (Patlolla et al., 2012), respectively, have been demonstrated with the GHS and SMT platforms, but global coverages do not exist at these resolutions.

Although most population estimation relies on census data, there are a handful of relevant examples of census-independent (“bottom-up”) approaches to mapping residential populations in data-poor environments. In one approach (Checchi et al., 2013), density estimates derived from literature and internet sources were used in conjunction with manual counts of structures from satellite imagery to estimate counts of displaced persons in eleven sites (a mixture of camps and urban neighborhoods) in Asia and Africa, and the largest estimation errors were seen where the density reports were scarce or unreliable, and/or where individual structures were difficult to discern from imagery. Another study (Hillson et al., 2014) used field surveys in Bo, Sierra Leone, to gather population observations and manual image interpretation to count buildings and measure their rooftops. An occupancy-based model (people per structure) was found to be more accurate than a rooftop area-based model, but the authors stressed the importance of practical considerations when choosing a density denominator. A third study (Stewart et al., 2016) estimated daytime and nighttime population using population density models derived from literature and internet sources and linked to specific facility types. Again, building footprints and classifications were identified manually from satellite and street-level imagery.

In this paper, we tackle the problem of unreliable and outdated census population counts through a bottom-up population mapping approach that couples semi-automated high-resolution settlement mapping with microcensus surveys, which are enumerations for sample zones within the settlement area, to estimate residential populations without relying on national census data. Our primary focus is on estimating the total residential population with high spatial precision, which can then serve as the denominator for estimating subpopulations when used in concert with known or estimated demographic, socioeconomic, or epidemiological rates (or, conversely, for estimating such rates in concert with observations of the numerators). We demonstrate subpopulation estimation by estimating the population of children under 5, a key demographic group for many health and development applications, including polio eradication. But this is just one possible application; the core of our approach has general applicability for any initiative aiming to accurately locate human settlements and estimate (sub-) populations in regions where census data are outdated or spatially imprecise.

2. Methods

2.1. Overview

Our approach to estimating residential population counts relies on three major components: a binary spatial layer of human-inhabited areas (the settlement layer), a categorical spatial layer of residential settlement types (the residential type layer), and a model of population density. The settlement layer and residential type layer are generated through remote sensing methods, while the population density model is specified using survey data from a microcensus.

To demonstrate and validate an approach to applying the population estimates toward the estimation of a subpopulation, we introduce a fourth component, a set of previously published demographic estimates (Alegana et al., 2015). We use the published estimates of the under-5 fraction of the population in conjunction with our population estimates to derive estimates of under-5 population counts for ten wards in Kano state, for which independent validation data are available. A graphical outline of the overall approach is shown in Fig. 1.

From the top-down modeling literature, we borrow the concept of estimating populations using land classes, but we adapt it to a census-independent setting. Our approach resembles that of Mennis and Hultgren (2006), but rather than selecting representative census