



Spatial and temporal changes in household structure locations using high-resolution satellite imagery for population assessment: an analysis in southern Zambia, 2006-2011

Timothy Shields,¹ Jessie Pinchoff,¹ Jailos Lubinda,² Harry Hamapumbu,² Kelly Searle,¹ Tamaki Kobayashi,¹ Philip E. Thuma,² William J. Moss,¹ Frank C. Curriero¹

¹*Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA;*

²*Macha Research Trust, Choma, Zambia*

Abstract

Satellite imagery is increasingly available at high spatial resolution and can be used for various purposes in public health research and

Correspondence: Timothy Shields, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, 615 N Wolfe Street, 21205 Baltimore, MD, USA.

Tel: +1.410.502.9077 - Fax: +1.410.955.0105.

E-mail: tshields@jhu.edu

Key words: Satellite imagery; Spatial epidemiology; GIS; Population targeting; Spatial statistics.

Contributions: TS assisted with analysis plan, derived and mapped all satellite imagery, and drafted the manuscript; JP assisted with data analysis and manuscript preparation; JL assisted with data management and mapping; HH ran the field team and all data collection; KS assisted with SaTScan; TK assisted with study design and preparation of manuscript; PET assisted with study design and coordination on site; WJM designed and coordinated the study and was involved with preparation of the manuscript; FCC designed the analysis plan and was involved in preparation of the manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: this work was supported by the Johns Hopkins Malaria Research Institute, the Bloomberg Family Foundation and the Division of Microbiology and Infectious Diseases, National Institutes of Allergy and Infectious Diseases, National Institutes of Health as part of the International Centers of Excellence for Malaria Research (U19 AI089680).

Acknowledgements: the researchers would like to thank the Macha Research Trust and community of Macha without which this research could not have been carried out.

Received for publication: 17 September 2015.

Revision received: 26 December 2015.

Accepted for publication: 10 January 2016.

©Copyright T. Shields *et al.*, 2016

Licensee PAGEPress, Italy

Geospatial Health 2016; 11:410

doi:10.4081/gh.2016.410

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

programme implementation. Comparing a census generated from two satellite images of the same region in rural southern Zambia obtained four and a half years apart identified patterns of household locations and change over time. The length of time that a satellite image-based census is accurate determines its utility. Households were enumerated manually from satellite images obtained in 2006 and 2011 of the same area. Spatial statistics were used to describe clustering, cluster detection, and spatial variation in the location of households. A total of 3821 household locations were enumerated in 2006 and 4256 in 2011, a net change of 435 houses (11.4% increase). Comparison of the images indicated that 971 (25.4%) structures were added and 536 (14.0%) removed. Further analysis suggested similar household clustering in the two images and no substantial difference in concentration of households across the study area. Cluster detection analysis identified a small area where significantly more household structures were removed than expected; however, the amount of change was of limited practical significance. These findings suggest that random sampling of households for study participation would not induce geographic bias if based on a 4.5-year-old image in this region. Application of spatial statistical methods provides insights into the population distribution changes between two time periods and can be helpful in assessing the accuracy of satellite imagery.

Introduction

Incorporating data from high-resolution satellite images and global positioning systems (GPS) into geographical information systems (GIS) has become increasingly useful, accurate, and widely available. Spatial resolution has dramatically increased to less than two meters for multispectral images and less than half a meter for panchromatic images. Availability has also increased, with a growing number of companies launching satellites for commercial use as well as some data freely available via web services such as Google Earth™ and Bing™ (Belward and Skoien, 2014). Data from satellites are spatially precise and spatial accuracy can be validated by GPS (Lowther *et al.*, 2009; Vazquez-Prokopec *et al.*, 2009; Checchi *et al.*, 2013).

High-resolution satellite images have diverse applications, such as the measurement of land use, population movement, change in civil infrastructure, conservation, monitoring of humanitarian emergencies, and the study of infectious diseases (Radke *et al.*, 2004; Dambach *et al.*, 2009; Schmidt and Kedir, 2009; Checchi *et al.*, 2013; Boyle *et al.*, 2014). In public health, the use of high-resolution satellite imagery has been identified as a cost-effective approach to develop disease surveillance systems, monitor disease trends, and document topographi-

cal changes that may influence disease transmission (Fernandez, 2008; Lefer *et al.*, 2008; Chang *et al.*, 2009; Kamadiou, 2009; Lowther *et al.*, 2009; Wei *et al.*, 2012; Soti *et al.*, 2013; Franke *et al.*, 2015; Nsoesie *et al.*, 2015). Another use for satellite imagery is in the selection of households for targeted interventions. For example, satellite imagery was used in Zambia to enumerate structures and select target areas to receive indoor residual spraying for malaria control (Franke *et al.*, 2015; Kamanga *et al.*, 2015).

Household enumeration based on high-resolution satellite imagery has been used to measure population changes in refugee camps, and has been identified as a practical method for generating a sampling frame for public health research in sub-Saharan Africa (Lowther *et al.*, 2009; Moss *et al.*, 2011; Wampler *et al.*, 2013; Escamilla *et al.*, 2014; Franke *et al.*, 2015). Satellite image-based census enumeration has also been employed to create population distribution maps that can be useful for many epidemiological calculations and studies, as well as for public health planning and targeting interventions (Chang *et al.*, 2009; Linard *et al.*, 2012; Wampler *et al.*, 2013; Kondo *et al.*, 2014). Unfortunately, existing census and demographic datasets for low-income countries, where disease burdens are commonly highest, are often based on outdated population enumeration data (Tatem *et al.*, 2007; Linard *et al.*, 2010, 2012). Recently, studies have validated the use of satellite imagery and GPS to provide sampling frames for ethnographic and public health surveys (Lowther *et al.*, 2009; Wampler *et al.*, 2013) and to estimate population size (Lowther *et al.*, 2009; Checchi *et al.*, 2013; Wampler *et al.*, 2013; Hillson *et al.*, 2014). However, the length of time a satellite image remains accurate and useful is unclear. Determining the accuracy of an image depends on the context, and varies based on the research question. For example, imagery utilized for epidemiological studies relying on household locations for survey implementation may be more temporally sensitive than studies determining and involving general land cover characteristics. Describing changes in the distribution of household structures visualised on satellite images is a novel application; previously, this technique has been restricted mainly to the description of refugee camps or areas of conflict (Galway *et al.*, 2012; Checchi *et al.*, 2013).

As a component of the Southern Africa International Centers of Excellence for Malaria Research (ICEMR), households are selected for enrolment into a prospective study of malaria transmission using simple random sampling from an enumerated list. Households in the sampling pool are identified and enumerated from a high-resolution satellite image and their coordinates confirmed by GPS at enrolment. For this sampling strategy to be effective, the coordinates of selected households must be accurate. Equally important is that the pool of enumerated households is accurately identified as collected field data are assumed to be representative of the target population. The temporal accuracy (*shelf-life*) of high-resolution satellite imagery was assessed by comparing images obtained in 2006 and 2011 of the study area in rural southern Zambia.

Materials and Methods

Study area

The catchment area of Macha Hospital in Choma District, Southern Province, Zambia is one of three sites of the ICEMR. The study site is a rural area approximately 575 km² at an average elevation of 1100 meters and consists of open savannah woodland with land clearings for subsistence agriculture (Moss *et al.*, 2011). All houses and non-residential structures are single story.

Geographical information system methodology

A satellite image task order was generated by DigitalGlobe Services, Inc (Denver, CO, USA) and a multispectral 2.4-m resolution image was acquired on 01/12/2006. This image was pan-sharpened to 0.62-meter resolution using the resolution-merge function. A second task order was generated by Apollo Mapping (Boulder, CO, USA) of the same study area for acquisition of a GeoEye-1 image obtained in mid-2011 with a 0.5-meter resolution. Six tiles from the imagery archives, collected between April and July 2011 (21/04/2011, 24/04/2011, 13/05/2011, 16/05/2011, and 18/07/2011) were added into a mosaic covering the study area. Orthorectification, using rational polynomial coefficients, was performed to improve spatial accuracy. Image processing was conducted in Erdas Imagine 2010 (Hexagon Geospatial, Norcross, GA, USA). Each image was imported into ArcGIS 9.2 (ESRI, Redlands, CA, USA). Visual inspection of the imagery was performed during the on-screen digitising process, during which structures of appropriate size and shape were identified as potential households. A typical household was typically recognised by a clearing of the natural brush with one or more domestic structures. Smaller structures, such as cooking houses or animal *kraals* might be present as well. A household was defined as one or more of these structures that function as a family unit. During this manual enumeration process, a map feature (point) was created for the centroid of each household. Comparison of the images allowed for the identification and coding of households that remained at the same location, were newly built, or were removed in the four and a half year period between the two images.

As an alternative to the manual enumeration process described above, household identification was originally attempted, without success, using automated feature extraction software. These software algorithms incorporate spatial context while classifying object-specific features specified by the user. However, in this study area and similar study areas in developing countries, the assortment of materials used for roofs (bush material, asbestos sheets, and corrugated metal) and walls (mud brick and concrete) impeded the ability to accurately and reliably discern houses. Additionally, our malaria data is collected and mapped at a household level, which, as stated, is often a collection of individual houses of varying number and geographic expanse.

Statistical analyses

Spatial statistics were used to assess clustering, cluster detection and spatial variation in household location between the 2006 and 2011 satellite images to describe and quantify changes in spatial patterns of households and to identify geographic areas of significant change.

Spatial clustering is the property that describes how tightly compact or dispersed a set of mapped locations are. The K-function, which estimates the expected number of other events within a range of distances of each event, was used to assess spatial clustering (Waller and Gotway, 2004). The K-function was estimated for both the 2006 and 2011 mapped household locations and the difference was plotted as a function of distance to assess change in the level of spatial clustering of household locations. Significant differences in spatial clustering were assessed using the Monte Carlo random labelling approach (Diggle, 2008). To complement the assessment of spatial clustering, spatial variation in the location of households was also explored. Spatial intensity, defined as the expected number of events per unit area, was estimated using the non-parametric kernel density approach and mapped to highlight spatial variation in the concentration of events (Waller *et al.*, 2004; Diggle, 2008). Spatial intensity was estimated for both the 2006 and 2011 mapped household locations. A map of the difference in spatial intensity between 2006 and 2011 was generated to show

changes in spatial variation of household locations between these two time periods. K-function and spatial intensity analysis were performed using the R Statistical Software with contributed *spatstat* package (Baddeley and Turner, 2005; R Statistical Software, 2013).

A cluster detection analysis was performed to assess clusters of significant change in the number households from 2006 to 2011. In comparison to the property of spatial clustering, a spatial cluster describes the local property of a subarea with a significant difference in the expected number of events. The existence of such a cluster may not be captured in the previously described analyses but could have profound effects on the sampling strategy and other related objectives that are based on enumerated satellite imagery. The study area was divided into 1-km grid cells. For each cell, the total number of newly added and removed households from 2006-2011, as well as the ratio of net change (difference in the added and removed houses) to the 2006 cell population, were determined. The cluster detection software SaTScan v9.4 (<http://www.statcan.org>) was used to search for clusters (contiguous sets of grid cells) with significantly high net change in household pop-

ulation from 2006 to 2011. The cluster detection was based on the SaTScan normal model to accommodate positive and negative net change and was performed controlling for proximity to roads (defined as the total length of all road segments in each grid cell). A tarred road was constructed in 2008 between the time points of the two images. Cluster detection analysis controlling for proximity to roads, a known driver of household settlement in this area, identifies clusters beyond what would have been explained by these features.

Results

A total of 3821 household structures were enumerated in 2006 and 4256 in 2011 (Table 1). Between 2006 and 2011, 971 (25.4%) structures were added and 536 (14.0%) structures removed (no longer present) (Table 1). Thus, by mid-2011, there was a net increase in 435 (11.4%) household structures from 2006. All enumerated household structures

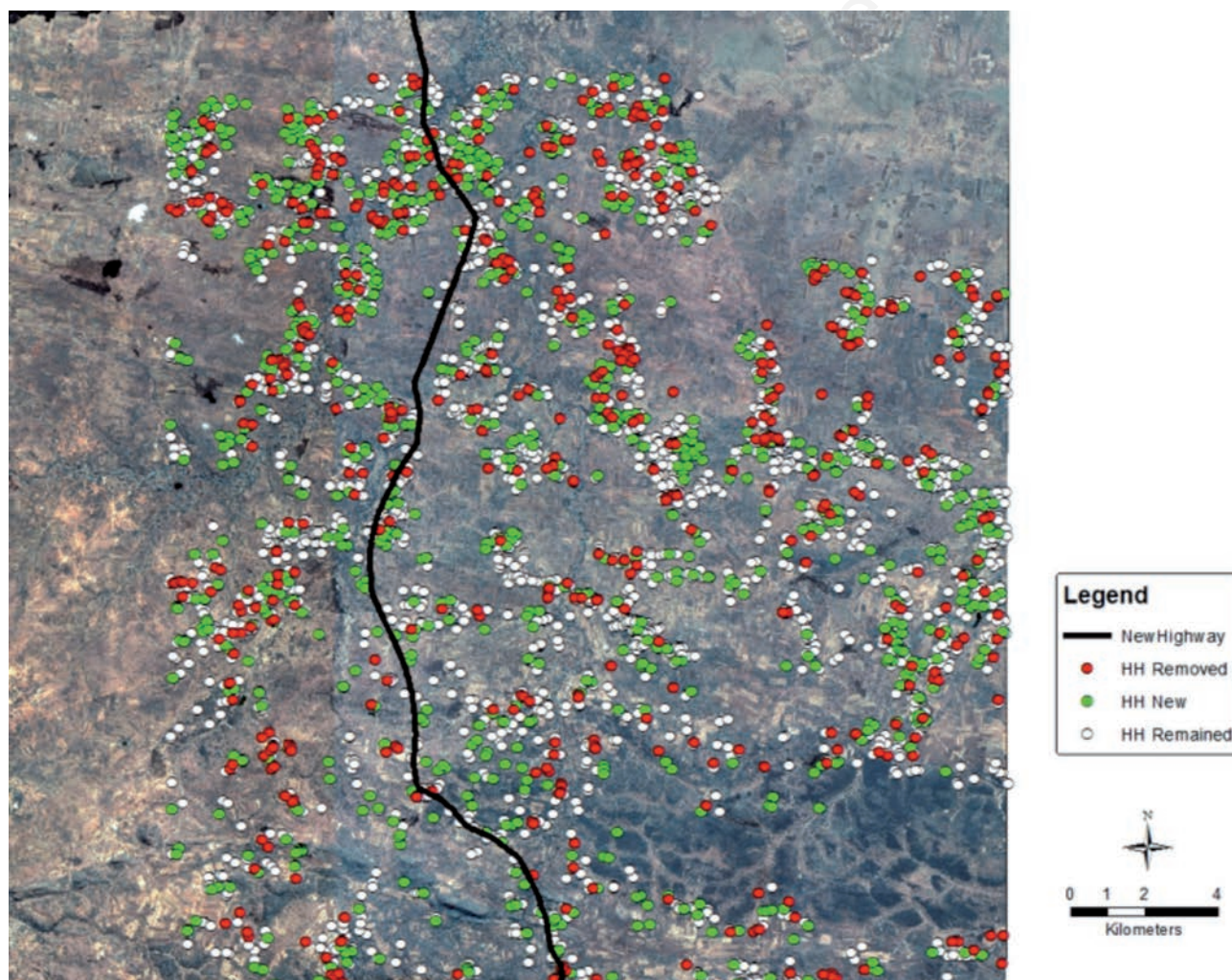


Figure 1. Change in households between the enumerated 2006 and 2011 satellite images for the study area in Southern Province, Zambia.

as well as the change (added and removed households) were mapped (Figure 1).

There was no significant difference in the level of spatial clustering for the 2006 household locations compared to the 2011 household locations. The difference in K-functions for 2006 to 2011 remained close to the horizontal zero line of no difference and did not approach statistical significance in either direction (Figure 2). Assessment of the intensity maps suggested the spatial variation in household concentrations were consistent from 2006 to 2011, although household density reached 32 houses per km² in 2011 compared to 27 per km² in 2006, reflecting the positive net change in households (Figure 3). The difference in intensity maps suggested that areas with the highest net change (both positive and negative) occurred where there were higher concentrations of households. An area of negative net change (more households removed than expected) appeared along the southern border of the northeast quadrant where there was moderate household density in both 2006 and 2011 (Figure 3).

The results of the SaTScan analysis identified one spatial cluster with a combined significantly negative net change (Figure 4). Within the cluster, significantly more household structures were removed between 2006 and 2011 than expected ($P < 0.001$). Although statistically significant, the weighted mean net change in the cluster was -0.23 houses compared to 0.15 houses outside the cluster.

Discussion

Satellite images depict the Earth's surface at a precise moment, providing a snapshot in time. Often missing from the literature, particularly in the field of public health, is an assessment of the degree of change over time and the potential impact such changes have on public health research, planning interventions, and population sampling. Assessment of change is also critical for longitudinal projects that involve planning, ongoing data collection, and outcome evaluations over space and time. Over a 4.5-year period, the number of households identified in a rural area of Zambia increased 11.4%; however, the household distribution patterns were maintained. These methodological approaches to examining changes in satellite imagery between two time periods can be used in other settings and for different research questions. The cost of acquiring new satellite imagery, although decreasing, remains an obstacle to their use in public health studies. Researchers have to determine whether existing archived imagery, which is significantly less expensive, is suitable for the research project and for how long a purchased image will remain useful. In public health studies, population movement is often a concern. Triggers such as changes in access and availability of transportation (e.g. road construction), new industrial developments, and changes in government policy can provide an indication that the population distribution in a given area may be changing. This analysis demonstrates how, with the use of spatial statistical techniques, these features can be incorporated into an assessment of change across multiple high-resolution satellite images of the same area. Identifying a net difference in the number of households between two time periods alone does not adequately describe the dynamics of household distribution. Further investigation highlighted that there were nearly twice as many households added as were removed. However, no significant change in the spatial distribution of household locations was identified in both large-scale spatial trends in the concentration of households and smaller scale spatial clustering of households. Although a statistically significant cluster of lower than expected net change in households was identified, the mag-

nitude of the difference was not deemed to be of practical significance for population sampling. This study incorporated time-intensive manual identification of households that was necessitated by the varying materials used in the construction of these houses and the need to identify groups of houses rather than individual structures. Automated feature extraction, including identification of houses, has been successfully utilised in other studies (Tullis and Jensen, 2003; Lo, 2007; Lowther *et al.*, 2009; Moss *et al.*, 2011; Wampler *et al.*, 2013; Escamilla *et al.*, 2014; Franke *et al.*, 2015; Kamanga *et al.*, 2015). Regardless of the method used to identify the map feature of interest the spatial statistical approach to understand the changes in imagery between time periods remains applicable. This study had some limitations. First, the assessment of household location change relied on observations from two images; no interim images were considered. Thus, longitudinal assessments at smaller temporal scales could not be determined. Second, the household enumeration process was based upon visual inspection of the images, potentially leading to the misclassification of non-residential structures as households or households as non-residential structures. Attempts were made to use automated feature

Table 1. Change with respect to the enumerated households for the 2006 and 2011 satellite imagery.

Household data	N	%
Households 2006	3821	na
Households 2011	4256	na
Households added between 2006 and 2011	971	25.4
Households removed between 2006 and 2011	536	14.0
Net increase	435	11.4

na, not available.

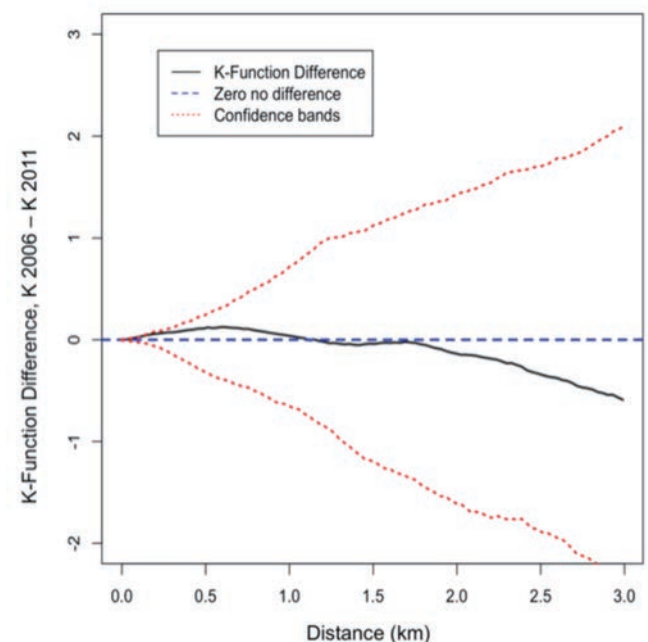


Figure 2. Difference in K functions comparing spatial clustering of enumerated household locations, 2006 to 2011.

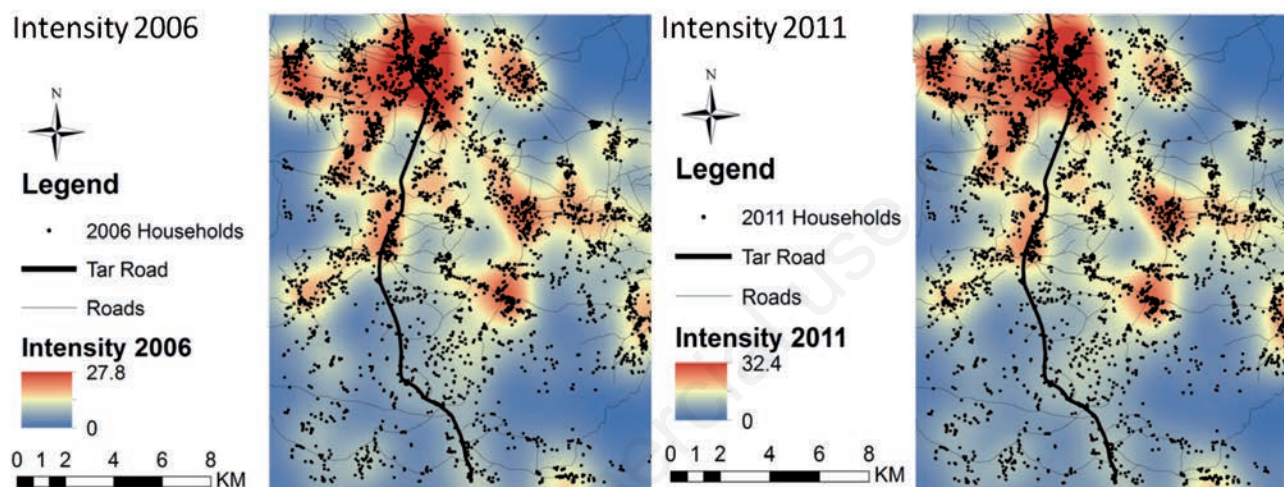
extraction software but none were able to account for the differing nature of the household materials. However, misclassified households were likely to be few or resulted in non-differential misclassification as the same methods were used for each image. Third, observations were based on enumerating household structures not actual people. While the number of household structures would likely be correlated with population size, on small spatial scales changes in population may not always be reflective of changes in household locations. Lastly, rural areas in southern Africa or in other developing countries may have more or

less household movement over time, thus limiting the generalisability of our findings. However, the methods used to assess changes in household structure patterns can be applied in different settings.

Conclusions

Satellite imagery is increasingly used for activities such as study planning, data collection, distribution of resources, or targeting of

A



B

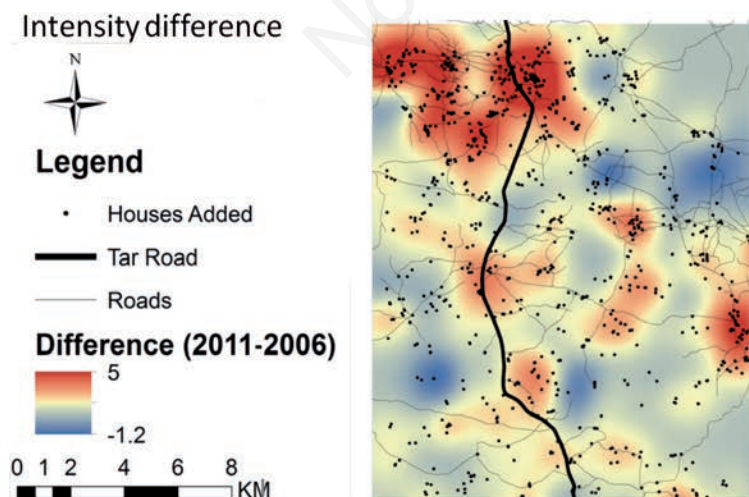


Figure 3. A) Mapped spatial intensity of enumerated household locations from the 2006 and 2011 satellite image; B) map of the difference in spatial intensity from 2006 to 2011.

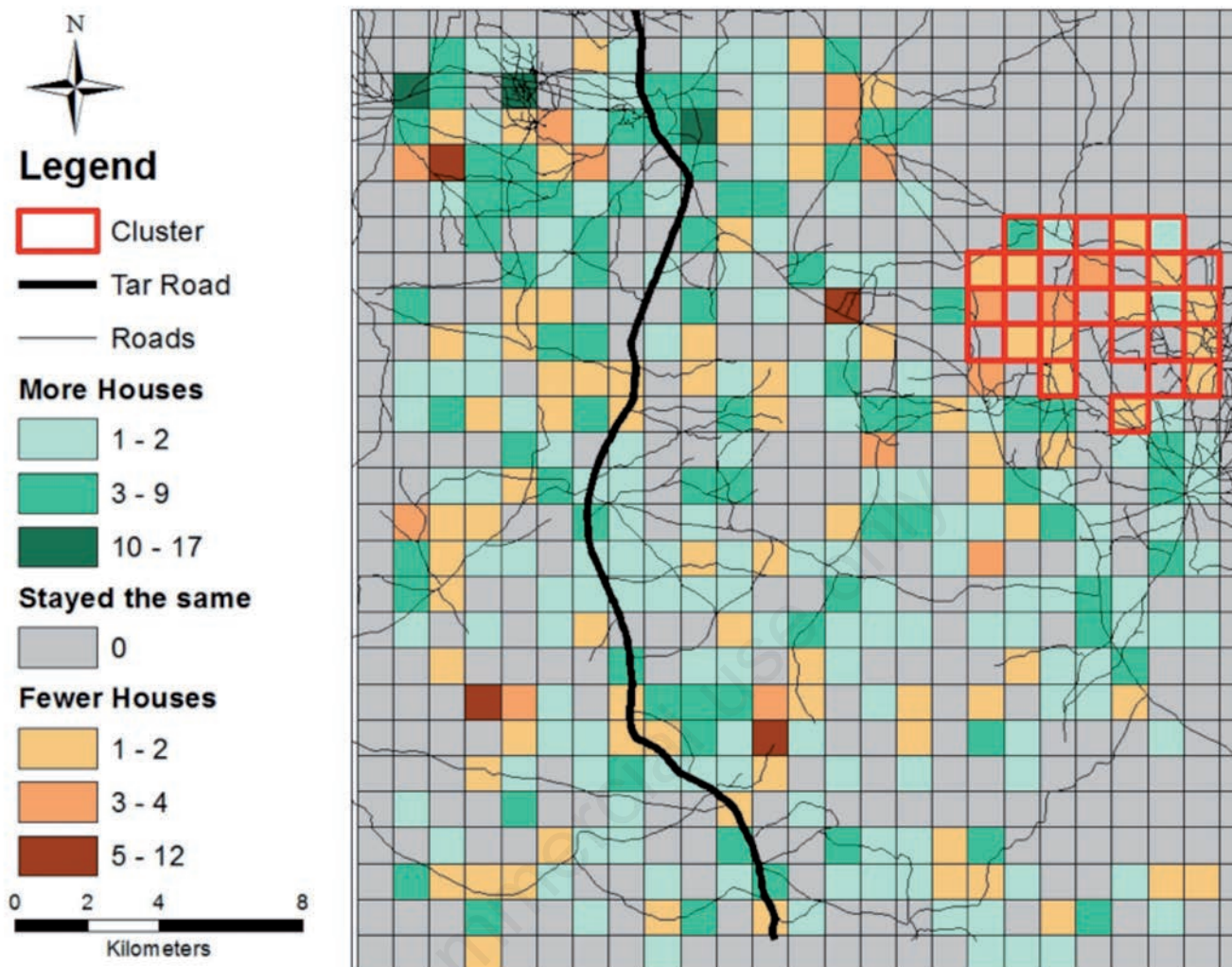


Figure 4. Net change in number of household structures within 1-km grid cells from 2006-2011, Choma District, Zambia.

activities. Understanding changes in the distribution of households over time is of importance to researchers relying on satellite imagery. Researchers should consider and evaluate the accuracy of satellite imagery as the time from acquisition to use increases.

References

- Baddeley A, Turner R, 2005. spatstat: an R package for analyzing spatial point patterns. *J Stat Softw* 12:1-42.
- Belward A, Skoien J, 2014. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS J Photogramm* 103:115-28.
- Boyle S, Kennedy C, Torres J, Colman K, Perez-Estigarribia P, 2014. High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PLoS One* 9:e86908.
- Chang A, Parrales M, Jimenez J, Sobieszczyk M, Hammer S, Copenhaver D, Kulkarni R, 2009. Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *Int J Health Geogr* 8:49.
- Checchi F, Stewart B, Palmer J, Grundy C, 2013. Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *Int J Health Geogr* 12:4.
- Dambach P, Sie A, Lacaux JP, Vignolles C, Machault V, Sauerborn R, 2009. Using high spatial resolution remote sensing for risk mapping of malaria occurrence in the Nouna district, Burkina Faso. *Glob Health Action* 2009:2.
- Diggle P, 2008. Statistical analysis of spatial patterns. Oxford University Press, New York, NY, USA.
- Escamilla V, Emch M, Dandalo L, Miller W, Martinson F, Hoffman I, 2014. Sampling at community level by using satellite imagery and geographical analysis. *B World Health Organ* 92:690-4.
- Fernandez I, 2008. Use of Google earth to facilitate GIS based decision support systems for arthropod-borne diseases. *Adv Dis Surveill* 4:91-2.
- Franke J, Gebreslasie M, Bauwens I, Deleu J, Siegert F, 2015. Earth observation in support of malaria control and epidemiology: MALAREO monitoring approaches. *Geospat Health* 10:335.
- Galway L, Bell N, Shatari A, Hagopian A, Burnham G, Flaxman A, Weiss



- W, Rajaratnam J, Takaro T, 2012. A two-stage cluster sampling method using gridded population data, a GIS, and Google EarthTM imagery in a population-based mortality survey in Iraq. *Int J Health Geogr* 11:12.
- Hillson R, Alejandre J, Jacobsen K, Ansumana R, Bockarie A, Bangura U, Lamin J, Malanoski A, Stenger D, 2014. Methods for determining the uncertainty of population estimates derived from satellite imagery and limited survey data: a case study of Bo city, Sierra Leone. *PLoS One* 9:e112241.
- Kamadieu R, 2009. Tracking the polio virus down the Congo River: a case study on the use of Google EarthTM in public health planning and mapping. *Int J Health Geogr* 8:4.
- Kamanga A, Renn S, Pollard D, Bridges DJ, Chirwa B, Pinchoff J, Larsen DA, Winters AM, 2015. Open-source satellite enumeration to map households: planning and targeting indoor residual spraying for malaria. *Malaria J* 14:435.
- Kondo M, Bream K, Barg F, Branas C, 2014. A random spatial sampling method in a rural developing nation. *BMC Public Health* 14:1-8.
- Lefer T, Anderson M, Fornari A, Lambert A, Fletcher J, Baquero M, 2008. Using Google Earth as an innovative tool for community mapping. *Publ Health Rep* 123:474-80.
- Linard C, Alegana V, Noor A, Snow R, Tatem A, 2010. A high resolution spatial population database of Somalia for disease risk mapping. *Int J Health Geogr* 9:45.
- Linard C, Gilbert M, Snow R, Noor A, Tatem A, 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 7:e31743.
- Lo CP, 2007. Automated population and dwelling unit estimation from high-resolution satellite images: a GIS approach. *Remote Sens* 16:17-34.
- Lowther S, Curriero F, Shields T, Ahmed S, Monze M, Moss W, 2009. Feasibility of satellite image-based sampling for a health survey among urban townships of Lusaka, Zambia. *Trop Med Int Health* 14:7-78.
- Moss W, Hamapumbu H, Kobayashi T, Shields T, Kamanga A, Clennon J, Mharakurwa S, Thuma P, Glass G, 2011. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community. *Malaria J* 10:163.
- Nsoesie E, Butler P, Ramakrishnan N, Mekaru S, Brownstein J, 2015. Monitoring disease trends using hospital traffic data from high resolution satellite imagery: a feasibility study. *Sci Rep* 5:9112.
- R Statistical Software, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Radke R, Andra S, Al-Kofahi O, Roysam B, 2004. Image change detection algorithms: a systematic survey. *IEEE T Image Process* 14:294-307.
- Schmidt E, Kedir M, 2009. Urbanization and spatial connectivity in Ethiopia: urban growth analysis using GIS. International Food Policy Research Institute, Addis Ababa, Ethiopia.
- Soti V, Chevalier V, Maura J, Begue A, LeLong C, Lancelot R, Thiongang Y, Tran A, 2013. Identifying landscape features associated with Rift Valley fever virus transmission, Ferlo region, Senegal, using very high spatial resolution satellite imagery. *Int J Health Geogr* 12:10.
- Tatem AJ, Noor AM, Hagen Cv, Gregorio AD, Hay SI, 2007. High resolution population maps for low income nations: combining land cover and census in East Africa. *PLoS One* 2:e1298.
- Tullis, JA, Jensen JR, 2003. Expert system house detection in high spatial resolution imagery using size, shape, and context. *Geocarto Intl* 18:5-15.
- Vazquez-Prokopec GM, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, Kochel TJ, Scott TW, Kitron U, 2009. Usefulness of commercially available GPS data-loggers for tracking human movement and exposure to Dengue virus. *Int J Health Geogr* 8:68.
- Waller LA, Gotway CA, 2004. Applied spatial statistics for public health data. Wiley, New York, NY, USA.
- Walsh SJ, Shao Y, Mena CF, McCleary AL, 2008. Integration of hyperion satellite data and a household social survey to characterize the causes and consequences of reforestation patterns in the Northern Ecuadorian Amazon. *Photogrammetric Eng Rem S* 74:725-35.
- Wampler P, Rediske R, Molla A, 2013. Using ArcMap, Google Earth, and global positioning systems to select and locate random households in rural Haiti. *Int J Health Geogr* 12:3.
- Wei L, Kun Y, Feng W, Xiao-Nong Z, Le-Ping S, Jian-Feng Z, Guo-Jing Y, Hang DR, Yong-Sheng J, 2012. A real-time platform for monitoring schistosomiasis transmission supported by Google Earth and a web-based geographical information system. *Geospat Health* 6:195-203.