Mapping the Environmental and Socioeconomic Coverage of the INDEPTH International Health and Demographic Surveillance System Network

Abstract

The International Network for the Demographic Evaluation of Populations and their Health (INDEPTH) has produced reliable longitudinal data about the lives of people in low- and middle-income countries (LMICs) through a global network of health and demographic surveillance system (HDSS) sites. Since reliable demographic data are scarce across many LMICs, we examine the environmental and socioeconomic (ES) similarities between existing HDSS sites and the rest of the LMICs. The HDSS sites were hierarchically grouped by the similarity of their ES conditions to quantify the ES variability between sites. The entire Africa and Asia region was classified to identify which regions were most similar to existing sites, based on available ES data. Results show that the current INDEPTH network architecture does a good job in representing ES conditions, but that great heterogeneities exist, even within individual countries. The results provide valuable information in determining the confidence with which relationships derived from present HDSS sites can be broadly extended to other areas, and to highlight areas where the new HDSS sites would improve significantly the ES coverage of the network.

Key Words

Health; remote sensing; Africa; Asia; demographic surveillance sites

Introduction

A health and demographic surveillance system (HDSS) undertakes ‘the longitudinal follow-up of well-defined entities or primary subjects (individuals, households and residential units) and all related demographic and health outcomes within a clearly circumscribed geographical area’ (INDEPTH, 2002). Ongoing cohort surveillance of health and demography in low- and middle-income countries (LMICs), conducted through a limited number of HDSS sites, provides valuable information on data-poor regions and reliable evidence for health policy design. Since the early 1960s, a collection of health research centers operating HDSSs have been established around specific intervention studies in LMICs, to monitor dynamics in populations,

1
2
3
4 including births, deaths, migrations and a range of other data (such as morbidity episodes,
5 pregnancies, marriages and economic activities), usually at 3-4 month intervals, in order to
6 quantify the long-term socioeconomic and demographic trends, as well as impacts of major
7 diseases and other external factors (Baiden et al., 2006; Tatem et al., 2006). In 1998, a number of
8 HDSS sites came together to found the International Network for the Demographic Evaluation of
9 Populations and their Health (INDEPTH) (Baiden et al., 2006; Sankoh and Byass, 2012).
10

11
12 One of the principal purposes of the INDEPTH network is to provide representative data
13 on a range of demographic and health indicators in low income regions of the world, where such
14 information is scarce (Sankoh and Byass, 2012). Further, INDEPTH aims to pool these site data
15 to provide information on the wide-area demographic patterns and their determinants. Morbidity,
16 mortality and health equity measures in Africa, however, are influenced strongly by both
17 socioeconomic conditions and the burden of infectious diseases (Murray and Lopez, 1997).
18 Therefore, to enable confident comparison between the individual HDSS sites across the
19 INDEPTH network and the effective extension of findings to other areas, the environmental and
20 socioeconomic characteristics of the sites should ideally be controlled for (Tatem et al., 2006).
21

22
23 A previous study examined the environmental coverage of the INDEPTH in rural Africa,
24 consisting of 21 HDSS sites at the time (Tatem et al., 2006). The analyses concluded that the
25 architecture of the INDEPTH rural network at the time provided a comprehensive coverage of
26 much of the wide range of climates and environments found across sub-Saharan Africa (SSA),
27 particularly across the southern Sahel and central tropical regions. This inferred, therefore, that
28 findings and relationships derived at those HDSS sites could be applied elsewhere within the
29 same environmental class with some confidence. Since that study, however, INDEPTH has
30 expanded its network, now incorporating 43 sites and covering three continents. Further, updated
31 and higher spatial resolution satellite-derived environmental datasets have become available, as
32 well as global spatial datasets representing a range of socioeconomic factors. This highlights the
33 need therefore to update, expand and improve previous assessments of the representativeness and
34 coverage of the INDEPTH network (Tatem et al., 2006) to account for these changes. Here we
35 undertake these analyses, focused on the new network, expanded geographical extent and new
36 global environmental and socioeconomic (ES) datasets. The goals of this study are two-fold: (i)
37 update the mapping of the environmental coverage of the newly expanded INDEPTH network,
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

going beyond solely Africa, and (ii) integrate more infrastructure and sociodemographic datasets that exist in a consistent format across Africa and Asia. The outputs focus on broad scales, and the typical conditions found at each site, and across the continents. We do not examine the demographic compositions and behaviors at each site, nor seasonal or inter-annual differences.

Methods

Data

At the time of writing, there were a total of 43 HDSS sites within the INDEPTH network: 30 in Africa, 12 in Asia and one in Oceania (www.indepth-network.org). Superimposed on the map in Figure 1 are the geographic locations of the 43 sites. Further descriptions are provided in supplemental information.

Data describing the environmental conditions across the region were derived from 1×1 km imagery from the MODerate-resolution Imaging Spectroradiometer (MODIS) sensors on-board NASA's Terra and Aqua satellites following the approaches outlined in Scharlemann et al. (2008). An imagery sequence from 2001 to 2008 was maximum value composited into monthly files and then subjected to Temporal Fourier Analysis (Rogers, 2000; Rogers et al., 2002). Of the outputs, the minimum, maximum and mean middle infrared radiation (MIR), land surface temperature (LST) and enhanced vegetation index (EVI) images were used here to provide broad quantitative descriptions of the environmental conditions experienced across Africa, Asia and Oceania.

EVI is an alternative version of the conventional Normalized Difference Vegetation Index (NDVI). NDVI has been regarded as a good indicator of soil moisture, rainfall, vegetation biomass and productivity, and neighborhood greenness (Rhew et al., 2011), and is the most commonly used vegetation index in epidemiological and health studies and associated with a large variety of health effects, such as diarrhea (Johnson et al., 2013), malaria prevalence (Gething et al., 2011), dengue transmission (Bhatt et al., 2013), land use, emerging infectious disease risk (Pigott et al., 2014) and coronary heart disease and stroke (Pereira et al., 2012), through direct and indirect relations to the vegetation cover, air quality, arthropod and animal

1
2
3
4 habitats, and physical activity. However, EVI has been proven to function with higher fidelity
5 through reducing saturation of the signal at high vegetation coverage and soil background effects
6 (Huete et al., 2002). LST is a measure of surface or “skin” temperature (Sims et al., 2008), and is
7 calculated by either split-window methods taking into account land surface emissivity or an
8 atmospheric correction method (Scharlemann et al., 2008). Due to more adequately representing
9 local temperature variability than ambient temperatures, LST has been widely associated with
10 heat-related health effects (Kestens et al., 2011), vector-borne infectious diseases (Bhatt et al.,
11 2013; Gething et al., 2011; Tatem et al., 2004), and drought impacts (Bohra-Mishra et al., 2014;
12 Karnieli et al., 2010). MIR, recorded by MODIS band 7, shows associations with the water
13 content, surface temperature and the structure of vegetation canopies (Boyd and Curran, 1998;
14 Tatem et al., 2006; Townshend and Justice, 2002), and thus as an indirect indication of vector-
15 borne disease risk (Hay, 2000; Hay et al., 2006).

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Monthly 1x1 km measurements of precipitation (PREC) averaged over the 1950-2000
period were obtained through WorldClim (Hijmans et al., 2005). The minimum, maximum and
mean precipitation, produced from all the values from 12 monthly layers, were used in this study
to represent typical patterns of precipitation across the region year round. The role of
precipitation on health and demographics has been identified in studies of water-borne diseases
(Curriero et al., 2001), malaria (Patz et al., 1998; Thomson et al., 2006), and migration (Garcia et
al., 2015; Henry et al., 2003), for example.

Three socioeconomic variables were used in addition to environmental variables to provide
an improved picture of conditions relating to health and demography in an area over simply
using environmental and climatic metrics: population density (POP), accessibility (ACCESS)
and economic output (ECON). Just as with the environmental variables, these do not capture the
full range of conditions and variabilities that exist in each area, but provide a broad set of metrics
that are available consistently across the two study continents. The Global Rural-Urban Mapping
Project (GRUMP) beta version dataset (Balk et al., 2006) was used to provide a representation of
population densities across the region studied. Population density and the effects of urbanization
have been shown to be related to a variety of demographic and health outcomes (Dye, 2008; Hay
et al., 2005). Accessibility, measured in terms of estimated travel time to major settlements, can
be a strong determinant of social, health and economic differences between regions (Linard et al.,

2012; Smith et al., 2013; World Bank, 2009). A global map of accessibility was obtained at 1km spatial resolution (Nelson, 2008), with grid cell values representing travel time in minutes to the nearest major settlement of population size 50,000 or more. A geographically based economic dataset, G-Econ 4.0 (Nordhaus et al., 2006) provided data on ‘gross cell product’, which estimates per-cell economic output values in U.S. dollars, as an indicator of economic productivity over 1×1 degree cells. The links between economic output, poverty and health are well known and widely studied and debated (e.g. (Feinstein, 1993)).

Pre-processing

Prior to analyses, all spatial datasets not originating from MODIS ($PREC_{max}$, $PREC_{mean}$, $PREC_{min}$, POP, ACCESS and ECON) were resampled and matched to the 1×1 km MODIS grids (EVI_{max} , EVI_{mean} , EVI_{min} , LST_{max} , LST_{mean} , LST_{min} , MIR_{max} , MIR_{mean} and MIR_{min}). Further, given that the focus of demographic and health-related studies are the populated areas of the world, all layers were masked to remove areas of population density less than 1 person/square km, as defined by GRUMP (Figure 1). Finally, the values in all spatial dataset layers were normalized to the same data range of 1 to 100 to ensure comparability.

Figure 1. The locations of HDSS sites and population density across Africa, Asia and Papua New Guinea of Oceania

ES signatures and distances

Buffers with a radius of 50km, centered on each INDEPTH HDSS site, were defined to represent the broad ES conditions experienced at HDSS sites relative to the rest of the regions under study. Buffers with different radii were tested (100km, 10km), but the choice was found to not influence results. Thus, for each HDSS site, minimum, maximum and mean MIR, LST, EVI, and precipitation, plus population density, accessibility and economic output values, were obtained, representing unique ES ‘signatures’ for each site (Tatem et al., 2006).

Given the 50km radius, five pairs of sites that were spatially proximate were merged as single sites for the analyses undertaken: Farafenni and West Kiang in Gambia (39.53km apart, 0.9963 correlation between signatures), Bandafassi and Mlomp in Senegal (3.76km, 0.9993), Ouagadougou and Sapone in Burkina Faso (49.24km, 0.9982), AMK2 and Matlab in Bangladesh (5.94km, 0.9997), and Filabavi and Dodalab in Vietnam (47.94km, 0.9998). Therefore, a total of 38 INDEPTH HDSS signatures were used in the analysis outlined in the remainder of this study.

1
2
3
4 To examine how similar the ES signature of each INDEPTH HDSS site was to the rest of
5 Africa, Asia and Papua New Guinea of Oceania, ES distances (Tatem et al., 2006) were
6 calculated using *ERDAS IMAGINE 2010* (Leica Geosystems) to produce a continuous measure
7 of ES similarity in 15-dimensional space between the signature of each grid square across the
8 continental regions and the signature of the site that it is most similar to. Thus, grid squares
9 showing large ES distance values were relatively dissimilar to the site that they were most alike,
10 while small distance values represented greater similarity to an existing site. The lowest distance
11 values were found at the INDEPTH HDSS sites, by definition. The overall output thus provided
12 a mapped representation of the ES coverage of the INDEPTH HDSS network. This
13 representation was also stratified by major Köppen climate region (Peel et al., 2007) to
14 separately highlight the coverage of the network in equatorial, arid, warm temperate and
15 cold/mountainous regions.
16
17
18
19
20
21
22
23
24
25
26
27
28
29

30 *Hierarchical clustering and classification*

31

32 To explore further the structure and coverage of the existing INDEPTH network in terms
33 of broad ES similarity, hierarchical clustering (Webb, 2003) of the site-specific signatures was
34 undertaken. In extending findings from individual INDEPTH network sites to other locations and
35 when wishing to summarize, data consistency in ES conditions is desirable, i.e. which sites are
36 relatively comparable and from which data could be compared or merged, as representative of a
37 wider region. Use of a dendrogram (tree diagram) enabled visualization and identification of
38 groupings of sites by ES similarity. The clustering process was undertaken in *R version 2.15.2* (R
39 Core Team 2012) on the basis of the similarities of signatures, producing a hierarchical tree
40 represented by the dendrogram (Tatem et al., 2006). The height of the dendrogram is measured
41 in Euclidean distance in multidimensional space, indicating how similar or dissimilar the branch
42 points (in this case, INDEPTH sites) are from each other at that height: the greater the height, the
43 greater the dissimilarity. No obvious changes in the structure of dendrogram were seen when
44 using either Mahalanobis or Euclidean distance measures.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59

60 **Results**

61
62
63
64
65

ES signatures of the INDEPTH HDSS sites

It is clear that, relative to the range of global variability, unsurprisingly most existing HDSS sites share some similar features, such as low economic output and population density, low PREC and MIR, moderately high EVI and high LST. Outliers exist though, for instance, higher MIR at the four HDSS sites in Burkina Faso and for Niakhar in Senegal. The Wosera site in Oceania had the lowest population density and poorest accessibility (as defined by the travel time measure used here), and second lowest economic output. The sites with relatively higher economic output and population density were almost all in Asia except for the Nairobi site. EVI in East Africa, Southern Africa, South Asia and Thailand tended to be overall much higher than others, while the volumes of precipitation at sites in South and Southeast Asia were higher than for those sites on the African continent. Details on the ES signatures of the INDEPTH sites are presented in supplemental information.

Cluster dendrogram and ES distances

Hierarchical clustering of site signatures produced the dendrogram shown in Figure 2, which provides visualization of the broad ES similarity between sites, with those close together and connected by short branches being most similar. Based on this, the 38 INDEPTH HDSS sites considered here were grouped into both 18, and a further 9 classes for better characterization (Figure 2), whereby those within the same dendrogram branch (and thus similar ES features) were assigned to the same class. More technical details are presented in supplemental information. In addition to being more similar in terms of ES conditions, grouped sites also reflected geographical clustering to some extent. For instance, all five sites in class *f* were in West Africa, four out of which were in Burkina Faso. Three out of four in class *o* are in East Africa, close to the juncture of Uganda, Kenya and Tanzania. Two out of five in class *p* were in West Africa and the other three in East Africa, while all three in class *q* were in East Africa as well. The sites in Asia tended to form smaller clusters, such as class *c*, *h* and *n*. The mean signatures of the 18 groups were normalized and their features are mapped in Figure 3 (more details in supplemental information), highlighting the ES features that characterize each group of sites (Figure 2) and the differences that exist between them.

Figure 4 shows the map of overall ES distance from the full set of INDEPTH HDSS sites with mean ES distance calculated (see supplemental information), and Figure 5 shows it broken down by climate zone. The figures demonstrate which areas are relatively most different environmentally and socioeconomically from the current network of sites, and how this representativeness breaks down by ecozone. It is clear that the least well represented regions include northern Sudan, eastern Turkey, western Iran, eastern Saudi Arabia, the western coastal plains in India, Central Asia, Russia, central Mongolia, northern China, North Korea, South Korea and Japan. Many of these are expected, and not countries for which the INDEPTH network is designed to be representative. However, the heterogeneity in distances within each country in Africa and Southeast Asia is indicative of great variations in the broad ES representativeness of the network. Further, more indepth analyses are presented in supplemental information.

Figure 2. Cluster dendrogram of the INDEPTH HDSS sites based on 15 ES variables. The sites are colored by class (Figure 3), with the lowercase letter on each branch identifying the class name. The within-class dissimilarities are smaller in the classes marked by lowercase (at the height of 50) than by uppercase letters (at the height of 100).

Figure 3. ES signature plot for each class of INDEPTH HDSS sites mapped in the dendrogram in Figure 2 (ACCESS: accessibility; ECON: economic output; POP: population; EVI_{mean} : mean Enhanced Vegetation Index; LST_{mean} : mean Land Surface Temperature; MIR_{mean} : mean Middle Infrared Reflectance; $PREC_{mean}$: mean precipitation), with each variable normalized to a 1-100 range, with 100 representing the largest values.

Figure 4. ES distances from the HDSS sites in the INDEPTH network in Africa and Asia. A large distance indicates a relatively different environment from that of the existing network of sites, while a small distance represents a relatively similar environment. Areas with population density of $<1 \text{ person/km}^2$ are masked in grey.

Figure 5. ES distances within five major climate regions. A large distance indicates a relatively different ES environment from that of existing HDSS sites, while a small distance represents a relatively similar ES environment.

Discussion

Since being founded in 1998, the INDEPTH network has grown to cover 20 countries in Africa, Asia and Oceania, following a population of 3.2 million to date. It continues to provide essential evidence for health policy making in low- and middle-income countries, filling important gaps globally in epidemiological surveillance (Sankoh and Byass, 2012). The data produced from INDEPTH sites has been used occasionally to represent conditions in wider, unsurveyed regions, where reporting is unreliable or non-existent, but how these data are extrapolated, and which sites are representative are rarely considered. A previous study attempted to tackle these issues (Tatem et al., 2006), but it was focused on Africa only, covered a smaller number of sites, examined solely environmental coverage and was based on satellite imagery that is now considered coarse and outdated. Here, each of these drawbacks have been overcome to provide a more detailed, contemporary, comprehensive and reliable assessment of the ES coverage of the existing INDEPTH network of HDSS sites.

Though it remains difficult to quantify what is ‘sufficient’, results show that the current INDEPTH network provides a relatively comprehensive coverage of the broad range of ES conditions that exist in the populated regions of Africa and Asia. Figure 4 highlights that the poorly represented ‘outlier’ regions are generally those for which the network was never intended to represent (e.g. Japan, Russia, Korea), though some key Sahelian regions still remain relatively unrepresented, similar to the findings of Tatem et al (2006). However, some of the key gaps identified in terms of environmental-only coverage in Tatem et al (2006) have been reduced through the addition of new sites and, importantly, despite subnational heterogeneities, the equatorial regions of the world that are among the poorest, are the most well represented by the network (Figure 5). Nevertheless, it is clear that uncertainties and sources of error remain to be tackled in future work.

We have endeavored here to utilize the most representative and recent data for each variable, available consistently across the continental study regions, to represent the contemporary ES conditions of the HDSS sites. Though some temporal mismatches among different variables involved in this study exist, the focus here was on providing broad measures of relative differences between locations, and capturing the typical conditions found at each site. Moreover, initial testing with some of the different datasets (MODIS and ECON for different

time periods) suggested no noticeable differences to results given the broad scales of analysis examined here. Future work will also focus on the exploration and construction of additional covariates to continue to work towards a more complete picture of the ES features of the regions studied. This will include datasets such as high resolution poverty maps (www.worldpop.org), variabilities in access to improved sanitation (Pullan et al., 2014) and subnational information on HIV prevalence (from e.g. <http://spatialdata.dhsprogram.com/home/>).

More efforts will be focused in future work on evaluating the demographic and health representativeness of the INDEPTH sites by comparing these sites with each other and to other locations where rich data exist in terms of factors such as age structures, mortality rates, and migration patterns of the populations. Through use of the data collected at the INDEPTH sites themselves, a much more detailed analysis of the sociodemographic similarities between sites could be undertaken, refining the similarity metrics calculated and the dendrogram analyses in Figure 2. This was beyond the focus of the current paper due to such rich data being unavailable beyond the sites themselves, but would be an informative study in terms of translating findings between sites. Obtaining these data in a standard and comparable format from all sites will however remain a challenging exercise, since not all sites make their data freely available. Other sources may be considered to complement the INDEPTH data though, such as mapped microdata from Terra Populus (<http://www.terrapop.org/>), or geolocated cluster data from the Demographic and Health Surveys (<http://spatialdata.dhsprogram.com/home/>).

The representativeness of the data produced from the HDSS themselves should be considered and accounted for. For instance, substantial heterogeneities in the population structures and proportion of migrants within the demographic surveillance area of the Nairobi HDSS exist (Emina et al., 2011), highlighting that the data from HDSS sites may over-represent permanent resident and under-represent the transient populations. Thus, in assessing the ‘representativeness’ of a site in terms of the surveillance data produced, these biases and heterogeneities should be considered (Collinson et al., 2007; White, 2009; Ye et al., 2012). While the INDEPTH network is one of the largest collections of surveillance sites providing ongoing demographic and health measurements, other sites producing similar data exist, such as the network of Institut Pasteur centres and other DSS sites that are not INDEPTH members. Extending the analysis presented here to include a wider set of surveillance sites will therefore be

a focus of future work, to produce an assessment of the comprehensiveness of current global surveillance, and where gaps exist. Additionally, while this work has considered temporal variability in terms of annual minimum, maximum and mean values of covariates, in reality, the seasonality in these factors can be of great importance to health and demographic outcomes, and this seasonality can be captured in covariates that are measured through space and time (e.g., from remotely sensed data sources). Environmental seasonality, in this sense, represents an important dimension of the environmental constraints that determine disease outcomes and needs to be captured as part of future analyses. Spatiotemporal methods and tools could be utilized and adapted to characterize seasonality in a few covariates for incorporation directly into the modeling (Alegana et al., 2013; Atkinson et al., 2012; Bakar and Sahu, 2013; Sahu et al., 2014). Seasonal parameters could be added to explain and adjust for seasonality, as has been undertaken previously (Sahu et al., 2014). Finally, while the analyses here provide a basic assessment of the similarities between existing sites and representativeness of the INDEPTH network in terms of available covariates, they do not fully account for the variation inherent within site catchments, or signatures across the two continents. For example, the Nairobi site in Kenya, located in two slum communities, does not represent Nairobi as a whole, nor are the precise conditions at these sites well represented by the ES data used here, due to data resolution limitations in terms of what exists presently for consistent measurements across continents. Here, for example, the Nairobi site is allocated a relatively high economic output value due to the coarse resolution of the global dataset that these estimates come from. Other urban sites also suffer from similar resolution issues, such as the Ouagadougou site that covers a number of peripheral slums, but being situated in a relatively high economic output area. Future work will therefore focus on the design and implementation of methods that can account for (i) site catchments and variability in covariate-defined conditions within them, (ii) site differences in terms of population numbers surveyed, and (iii) the range of variability in covariate values that make up the signature of each site. Here, hierarchical geostatistical models using the Bayesian paradigm could be utilised, making use of existing open software packages, e.g. spBayes (Finley et al., 2007) and INLA (Rue et al., 2009). This framework would enable rigorous accounting for the variability observed in site catchments, populations, health outcomes and signatures and propagate this variation through to the dendrogram and distance maps, enabling a more complete assessment of the uncertainties inherent in the analyses.

The above future work steps would ultimately lead to development of a flexible tool for providing general guidance for the positioning of new surveillance sites to ensure that a wide range of populations are well-represented. Though INDEPTH do not themselves establish new sites, other organisations continue to work towards better global surveillance (e.g. <http://www.gatesfoundation.org/Media-Center/Press-Releases/2015/05/Child-Health-and-Mortality-Prevention-Surveillance-Network>). The work presented here provides quantitative outputs that provide valuable guidance on where the existing coverage (in terms of available covariates) of surveillance sites is relatively good or poor, but further steps are required if the positioning of new sites is to be optimized or the required number of sites is to be determined based on pre-defined criteria. Here, approaches developed for the optimal design of spatial sampling strategies could be adapted (e.g., (Neal et al., 2012)). Optimum spatial sampling design has great potential in this context, but has seen limited application. The positioning of multiple new sites presents methodological challenges due to the interactions that each new site will have on prior and subsequent sites and, thus, the need to consider jointly their ability to optimally improve network coverage and representativeness of populations.

As the World transitions from the Millennium Development Goals (MDGs) era to the setting of new Sustainable Development Goals (SDGs), the monitoring of progress towards completing both sets of goals is as important as ever. The INDEPTH network, as well as other sets of longitudinal surveillance sites, is central to providing the evidence bases required for such progress tracking in countries where national surveillance systems are weak. Ensuring that such surveillance is as representative as possible of the range of environmental conditions and socioeconomic variabilities found across the regions of interest is therefore vital to accurate monitoring. Here we have outlined a framework and set of results for assessing this coverage across continents and providing quantitative evidence for guiding extrapolation and interpretation of surveillance outputs across broad areas.

References

- Alegana, V.A., Atkinson, P.M., Wright, J.A., Kamwi, R., Uusiku, P., Katokele, S., Snow, R.W., Noor, A.M., 2013. Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial-temporal models. *Spat Spatiotemporal Epidemiol* 7.
- Atkinson, P.M., Jeganathan, C., Dash, J., Atzberger, C., 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *REMOTE SENSING OF ENVIRONMENT* 123, 400-417.
- Baiden, F., Hodgson, A., Binka, F.N., 2006. Demographic surveillance sites and emerging challenges in international health. *Bull World Health Organ* 84, 163-163.
- Bakar, K.S., Sahu, S.K., 2013. spTimer: Spatio-Temporal Bayesian Modelling Using R. *Journal of Statistical Software* (in press).
- Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining global population distribution: methods, applications and data. *Adv Parasitol* 62, 119-156.
- Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., Myers, M.F., George, D.B., Jaenisch, T., Wint, G.R., Simmons, C.P., Scott, T.W., Farrar, J.J., Hay, S.I., 2013. The global distribution and burden of dengue. *Nature* 496, 504-507.
- Bohra-Mishra, P., Oppenheimer, M., Hsiang, S.M., 2014. Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proc Natl Acad Sci U S A* 111, 9780-9785.
- Boyd, D.S., Curran, P.J., 1998. Using remote sensing to reduce uncertainties in the global carbon budget: the potential of radiation acquired in middle infrared wavelengths. *Remote Sensing Reviews* 16, 293-327.
- Collinson, M.A., Tollman, S.M., Kahn, K., 2007. Migration, settlement change and health in post-apartheid South Africa: triangulating health and demographic surveillance with national census data. *Scand J Public Health Suppl* 69, 77-84.
- Curriero, F.C., Patz, J.A., Rose, J.B., Lele, S., 2001. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *Am J Public Health* 91, 1194-1199.
- Dye, C., 2008. Health and urban living. *Science* 319, 766-769.
- Emina, J., Beguy, D., Zulu, E.M., Ezech, A.C., Muindi, K., Elung'ata, P., Otsola, J.K., Ye, Y., 2011. Monitoring of health and demographic outcomes in poor urban settlements: evidence from the Nairobi Urban Health and Demographic Surveillance System. *J Urban Health* 88 Suppl 2, S200-218.
- Feinstein, J.S., 1993. The relationship between socioeconomic status and health: a review of the literature. *Milbank Q* 71, 279-322.
- Finley, A.O., Banerjee, S., Carlin, B.P., 2007. spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software* 19, 1-24.
- Garcia, A.J., Pindolia, D.K., Lopiano, K.K., Tatem, A.J., . 2015. Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies* 3, 89-110.

- Gething, P.W., Patil, A.P., Smith, D.L., Guerra, C.A., Elyazar, I.R., Johnston, G.L., Tatem, A.J., Hay, S.I., 2011. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar J* 10, 378.
- Hay, S.I., 2000. An overview of remote sensing and geodesy for epidemiology and public health application. *Adv Parasitol* 47, 1-35.
- Hay, S.I., Guerra, C.A., Tatem, A.J., Atkinson, P.M., Snow, R.W., 2005. Urbanization, malaria transmission and disease burden in Africa. *Nat Rev Microbiol* 3, 81-90.
- Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J., Rogers, D.J., 2006. Global environmental data for mapping infectious disease distribution. *Adv Parasitol* 62, 37-77.
- Henry, S., Boyle, P., Lambin, E.F., 2003. Modelling inter-provincial migration in Burkina Faso, West Africa: the role of socio-demographic and environmental factors. *Applied Geography* 23, 115-136.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology* 25, 1965-1978.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *REMOTE SENSING OF ENVIRONMENT* 83, 195-213.
- INDEPTH, 2002. Population and health in developing countries, Population, Health and Survival at INDEPTH Sites. International Development Research Centre, Ottawa.
- Johnson, K.B., Jacob, A., Brown, M.E., 2013. Forest cover associated with improved child health and nutrition: evidence from the Malawi Demographic and Health Survey and satellite data. *Glob Health Sci Pract* 1, 237-248.
- Karnieli, A., Agam, N., Pinker, R.T., Anderson, M., Imhoff, M.L., Gutman, G.G., Panov, N., Goldberg, A., 2010. Use of NDVI and land surface temperature for drought assessment: merits and limitations. *Journal of Climate* 23, 618-633.
- Kestens, Y., Brand, A., Fournier, M., Goudreau, S., Kosatsky, T., Maloley, M., Smargiassi, A., 2011. Modelling the variation of land surface temperature as determinant of risk of heat-related health events. *Int J Health Geogr* 10, 7.
- Linard, C., Gilbert, M., Snow, R.W., Noor, A.M., Tatem, A.J., 2012. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 7, e31743.
- Murray, C.J., Lopez, A.D., 1997. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 349, 1269-1276.
- Neal, J.C., Atkinson, P.M., Hutton, C.W., 2012. Adaptive space-time sampling with wireless sensor nodes for flood forecasting. *Journal of Hydrology* 414, 136-147.
- Nelson, A., 2008. Estimated travel time to the nearest city of 50,000 or more people in year 2000, in: Commission, G.E.M.U.-J.R.C.o.t.E. (Ed.), Ispra, Italy.
- Nordhaus, W., Azam, Q., Corderi, D., Hood, K., Victor, N.M., Mohammed, M., Miltner, A., Weiss, J., 2006. The G-Econ database on gridded output: methods and data. Yale University.

- Patz, J.A., Strzepek, K., Lele, S., Hedden, M., Greene, S., Noden, B., Hay, S.I., Kalkstein, L., Beier, J.C., 1998. Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya. *Trop Med Int Health* 3, 818-827.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions* 4, 439-473.
- Pereira, G., Foster, S., Martin, K., Christian, H., Boruff, B.J., Knuiman, M., Giles-Corti, B., 2012. The association between neighborhood greenness and cardiovascular disease: an observational study. *BMC Public Health* 12, 466.
- Pigott, D.M., Golding, N., Mylne, A., Huang, Z., Henry, A.J., Weiss, D.J., Brady, O.J., Kraemer, M.U., Smith, D.L., Moyes, C.L., Bhatt, S., Gething, P.W., Horby, P.W., Bogoch, II, Brownstein, J.S., Mekaru, S.R., Tatem, A.J., Khan, K., Hay, S.I., 2014. Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* 3, e04395.
- Pullan, R.L., Freeman, M.C., Gething, P.W., Brooker, S.J., 2014. Geographical inequalities in use of improved drinking water supply and sanitation across Sub-Saharan Africa: mapping and spatial analysis of cross-sectional survey data. *PLoS medicine* 11, e1001626.
- Rhew, I.C., Vander Stoep, A., Kearney, A., Smith, N.L., Dunbar, M.D., 2011. Validation of the normalized difference vegetation index as a measure of neighborhood greenness. *Annals of epidemiology* 21, 946-952.
- Rogers, D.J., 2000. Satellites, space, time and the African trypanosomiases. *Adv Parasitol* 47, 129-171.
- Rogers, D.J., Randolph, S.E., Snow, R.W., Hay, S.I., 2002. Satellite imagery in the study and forecast of malaria. *Nature* 415, 710-715.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* 71, 319-392.
- Sahu, S.K., Baffour, B., Harper, P.R., Minty, J.H., Saran, C., 2014. A hierarchical Bayesian model for improving short-term forecasting of hospital demand by including meteorological information. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 177, 39-61.
- Sankoh, O., Byass, P., 2012. The INDEPTH Network: filling vital gaps in global epidemiology. *Int J Epidemiol* 41, 579-588.
- Scharlemann, J.P., Benz, D., Hay, S.I., Purse, B.V., Tatem, A.J., Wint, G.W., Rogers, D.J., 2008. Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. *PLoS One* 3, e1408.
- Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Bolstad, P.V., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L., Monson, R.K., Oechel, W.C., Schmid, H.P., Wofsy, S.C., Xu, L., 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *REMOTE SENSING OF ENVIRONMENT* 112, 1633-1646.
- Smith, D.L., Cohen, J.M., Chiyaka, C., Johnston, G., Gething, P.W., Gosling, R., Buckee, C.O., Laxminarayan, R., Hay, S.I., Tatem, A.J., 2013. A sticky situation: the unexpected stability of malaria elimination. *Philos Trans R Soc Lond B Biol Sci* 368, 20120145.

- 1
2
3
4 Tatem, A.J., Goetz, S.J., Hay, S.I., 2004. Terra and Aqua: new data for epidemiology and public
5 health. *Int J Appl Earth Obs Geoinf* 6, 33-46.
6
7 Tatem, A.J., Snow, R.W., Hay, S.I., 2006. Mapping the environmental coverage of the
8 INDEPTH demographic surveillance system network in rural Africa. *Trop Med Int Health* 11,
9 1318-1326.
10
11 Thomson, M.C., Doblas-Reyes, F.J., Mason, S.J., Hagedorn, R., Connor, S.J., Phindela, T.,
12 Morse, A.P., Palmer, T.N., 2006. Malaria early warnings based on seasonal climate forecasts
13 from multi-model ensembles. *Nature* 439, 576-579.
14
15 Townshend, J.R., Justice, C.O., 2002. Towards operational monitoring of terrestrial systems by
16 moderate-resolution remote sensing. *REMOTE SENSING OF ENVIRONMENT* 83, 351-359.
17
18 Webb, A.R., 2003. Statistical pattern recognition. John Wiley & Sons.
19
20 White, M.J., 2009. Migration and Demographic Surveillance: An overview of opportunities and
21 challenges, in: al., M.C.e. (Ed.), *Households Unfolding: INDEPTH Network Perspectives on the*
22 *Dynamics of Migration, Health and Livelihoods*. Ashgate.
23
24 World Bank, 2009. *World Development Report 2009: Reshaping Economic Geography*. World
25 Bank, Washington DC, p. 383.
26
27 Ye, Y., Wamukoya, M., Ezech, A., Emina, J.B., Sankoh, O., 2012. Health and demographic
28 surveillance systems: a step towards full civil registration and vital statistics system in sub-
29 Sahara Africa? *BMC Public Health* 12, 741.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Mapping the Environmental and Socioeconomic Coverage of the INDEPTH International Health and Demographic Surveillance System Network

Peng Jia¹, Osman Sankoh² and Andrew J. Tatem^{3,4,5*}

1 Department of Epidemiology and Environmental Health, University at Buffalo, Buffalo, NY, USA

2 INDEPTH, Accra, Ghana

3 Department of Geography and Environment, University of Southampton, Highfield, Southampton, UK

4 Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

5 Flowminder Foundation, Stockholm, Sweden.

*Corresponding author

Author's contribution

AJT and PJ designed the study. AJT, PJ and OS acquired the data. PJ led the analysis and interpretation of all data, prepared all drafts of the paper, and incorporated comments by coauthors. AJT made significant contribution to the interpretation of data and editing of the paper. PJ, OS, and AJT all approved the final version.

Acknowledgements

AJT is supported by funding from NIH/NIAID (U19AI089674), the Bill & Melinda Gates Foundation (OPP1106427, 1032350), and the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. This work forms part of the WorldPop Project (www.worldpop.org.uk) and Flowminder (www.flowminder.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.