

## LINKING SATELLITE REMOTE SENSING BASED ENVIRONMENTAL PREDICTORS TO DISEASE: AN APPLICATION TO THE SPATIOTEMPORAL MODELLING OF SCHISTOSOMIASIS IN GHANA

M. Wrable<sup>a,\*</sup>, A. Liss<sup>a,\*</sup>, A. Kulinkina<sup>a,\*</sup>, M. Koch<sup>b</sup>, N. K. Biritwum<sup>c</sup>, A. Ofosu<sup>d</sup>, K. C. Kosinski<sup>e</sup>, D. M. Gute<sup>a</sup>, E. N. Naumova<sup>f,\*</sup>

<sup>a</sup> Dept. of Civil and Environmental Engineering, Tufts University, 200 College Ave, Medford, USA  
– (madeline.wrable, alexander.liss, alexandra.kulinkina, david.gute)@tufts.edu

<sup>b</sup> Center for Remote Sensing, Boston University, Boston, USA – mkoch@bu.edu

<sup>c</sup> Neglected Tropical Disease Control Program, Ghana Health Service, Accra, Ghana – nanakwadwo.biritwum@ghsmai.org

<sup>d</sup> Policy, Planning, Monitoring, and Evaluation Division, Ghana Health Service, Accra, Ghana – anthony.ofosu@ghsmai.org

<sup>e</sup> Dept. of Community Health, Tufts University, 574 Boston Avenue, Medford, USA – karen.kosinski@tufts.edu

<sup>f</sup> Friedman School of Nutrition Science and Policy, Tufts University, 150 Harrison Ave, Boston, USA – elena.naumova@tufts.edu

### Commission VIII, WG VIII/2

**KEY WORDS:** Remote Sensing, Environmental Variables, Climate Classification, Spatiotemporal Analysis, Schistosomiasis Risk

### ABSTRACT:

90% of the worldwide schistosomiasis burden falls on sub-Saharan Africa. Control efforts are often based on infrequent, small-scale health surveys, which are expensive and logistically difficult to conduct. Use of satellite imagery to predictively model infectious disease transmission has great potential for public health applications. Transmission of schistosomiasis requires specific environmental conditions to sustain freshwater snails, however has unknown seasonality, and is difficult to study due to a long lag between infection and clinical symptoms. To overcome this, we employed a comprehensive 8-year time-series built from remote sensing feeds. The purely environmental predictor variables: accumulated precipitation, land surface temperature, vegetative growth indices, and climate zones created from a novel climate regionalization technique, were regressed against 8 years of national surveillance data in Ghana. All data were aggregated temporally into monthly observations, and spatially at the level of administrative districts. The result of an initial mixed effects model had 41% explained variance overall. Stratification by climate zone brought the  $R^2$  as high as 50% for major zones and as high as 59% for minor zones. This can lead to a predictive risk model used to develop a decision support framework to design treatment schemes and direct scarce resources to areas with the highest risk of infection. This framework can be applied to diseases sensitive to climate or to locations where remote sensing would be better suited than health surveys.

### 1. INTRODUCTION

Schistosomiasis is a debilitating infection acquired from contact with infested water bodies and caused by parasitic blood flukes of the genus *Schistosoma*. It infects over 200 million people worldwide, and around 779 million are at risk (Steinmann et al., 2006; Utzinger et al., 2009; Gray et al., 2010). Schistosomiasis is endemic to parts of South America, the Middle East, Asia, and Africa. Sub-Saharan Africa contributes 90% of cases worldwide (Chitsulo et al., 2000; Gryseels et al., 2006; Hurlimann et al., 2011) and Ghana has one of the highest prevalence rates of infections (WHO, 2010). In Ghana 30-35% of the population require preventative chemotherapy, however only 2–8% of the population receives it annually (WHO, 2010). Due to limited disease surveillance, control efforts are often based on infrequent, small-scale health surveys, which are expensive and logistically difficult to conduct.

Favorable snail habitat can serve as a proxy for potential schistosomiasis transmission, because snails must be present for the parasite to develop into a stage infective to humans (Sturrock et al., 2001). For disease to occur there must be

interaction between the parasite, snail, and human. The parasite and snail factors are primarily environmental (see Table 1). Thus, environmental variables that dictate snail habitat can be used to predict disease transmission. Furthermore, these parameters can be accessed by remote sensing (RS) technology (Simoonga et al., 2009). Specific RS data are publically available at no or limited cost, and since this technology can cover a large spatial and temporal scale, it is ideal for use in low-income countries lacking the resources to perform expensive ground-based studies (Gryseels et al., 2006). Recent review papers show that the most commonly used environmental variables are temperature, vegetation, rainfall, water chemistry, distance to waterbodies, and elevation (Simonga et al., 2009; Walz et al., 2015a). Aggregation over ecological zones as opposed to administrative units has been recommended in the literature, and is shown to be valuable in predicting the occurrence of schistosomiasis (Simoonga et al., 2008; Grosse, 1993; Walz et al., 2015b; Walz et al., 2015c).

\* Corresponding authors

| Parasite      | Snail                  | Human               |
|---------------|------------------------|---------------------|
| Temperature   | Water temperature      | Water contact       |
| Water flow    | Water flow/level/depth | Health, Controls    |
| Sunlight      | Sunlight, Vegetation   | Age, Gender         |
| Species       | Species                | Class/Pay, Job      |
| Predators     | Predators              | Ethnicity, Religion |
| Pathogenicity | Rainfall               | Residence, Travel   |

Table 1: Parasite, snail, and human factors affecting schistosomiasis transmission (adapted from Walz et al., 2015a).

In order to determine predictive properties of RS-based environmental parameters, we examined monthly records of reported schistosomiasis during 2008-2015 and their associations, with time-series built from remote sensing feeds at the district level in Ghana.

## 2. DATA

The environmental variables used in this work came entirely from publically available remote sensing data (Table 2). The environmental variables were selected based on their repeated use in the literature (Bavia et al., 2001; Simoonga et al 2009; Walz et. al 2015a), and for comparison to the predominant climate classification system for the past 100 years. The Köppen-Geiger (KG) climate classification system is based on the assumption that vegetation is the best proxy for climate, and that temperature and precipitation are the best proxies for vegetation (Kottek et al., 2006). The variables were projected to WGS-84/UTM-30N datum prior to aggregation to mean values per district, using zonal statistics in the ArcGIS 10.3 software. Where resampling was required, the cubic convolution technique was used because it more realistically reflected the smooth transitions of environmental data across terrain.

| Parameter         | Data Source        | Product      | Temporal Resolution         | Spatial Resolution |
|-------------------|--------------------|--------------|-----------------------------|--------------------|
| LST <sup>1</sup>  | MODIS <sup>5</sup> | MOD/MYD-11A2 | calculated 8-day composites | 1 km <sup>2</sup>  |
| NDVI <sup>2</sup> | MODIS <sup>5</sup> | MOD/MYD-13A2 | calculated 8-day composites | 1 km <sup>2</sup>  |
| LKNr <sup>3</sup> | MODIS <sup>5</sup> | MYD-13Q1     | 15-year aggregates          | Regional           |
| AP <sup>4</sup>   | TRMM <sup>6</sup>  | 3B43 v7      | Monthly                     | 28 km <sup>2</sup> |
| Rate              | GHS <sup>7</sup>   | Counts       | Monthly                     | District           |

<sup>1</sup> Land surface temperature (LST)

<sup>2</sup> Normalized Difference Vegetation Index (NDVI)

<sup>3</sup> LKN-regions (LKNr)

<sup>4</sup> Accumulated Precipitation (AP)

<sup>5</sup> Moderate Resolution Imaging Spectroradiometer (MODIS)

<sup>6</sup> Tropical Rainfall Monitoring Mission (TRMM)

<sup>7</sup> Ghana Health Service (GHS)

Table 2: Data sources and resolution of environmental parameters and the health outcome.

Ghana Health Service (GHS) provided monthly counts of schistosomiasis reported to the national surveillance system from January 2008 through December 2015. The records are likely to reflect the process of adult patients seeking medical treatment for schistosomiasis-related symptoms across 216 districts. The case counts were converted into monthly

prevalence rates, using district populations obtained from the 2010 Census and with intra-censal population growth projections (GSS, 2016). Rates were calculated per 10,000 people.

The original case counts contained “blanks” (47% of all data points), which might be interpreted as a missing data points or as “true” lack of reporting for a given month in a given location. In the current study “blanks” were replaced with zeros. To correct for right-skewed distribution, the rates were transformed by natural logarithm and adjusted by the lowest observed disease rate for a non-zero count of 0.00506.

## 3. METHODS

### 3.1 LKN regionalization

LKN (LKNr) regions were created following the methodology of our recently proposed automated climate regionalization method, LKN-regionalization, which are based on k-means clustering algorithm over time-space (Liss et al. 2015). The LKN method consists of the data limiting step (L-step) to reduce dimensionality by applying principal component analysis, a classification step (K-step) to produce hierarchical candidate regions using k-means unsupervised classification algorithm, and a nomination step (N-step) to determine the number of candidate climate regions using cluster validity indexes. Using a comprehensive set of multiple satellite data streams arranged as time series, the method is capable of defining climate regions over large spatial extents. This is essential for large-scale epidemiological studies to account for geographic heterogeneity.

MODIS Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) for 15 years (2000-2015) were downloaded from the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (LPDAAC-NASA 2000-2015). The data was mosaicked and re-projected so that it covered the entire extent of our study region (latitude and longitude of the upper left and lower right corners of the bounding box were: UL (11.625, -3.625), LR (4.375, 1.375)). Each of the two EOS satellites, Aqua and Terra, produced composites on overlapping 16 days schedules that we combined and constructed 8 day temporal resolution composites. The data was aggregated in a layered space-time series. Using a normalized vegetation index allowed us to reduce or eliminate the effects of annually changing lighting conditions, thin clouds, atmospheric and anisotropic distortions.

Water bodies were masked to avoid misclassification, because the water reflectance pattern differs significantly from almost any other land surface material by absorbing most of the incoming radiation. The imagery for each of the 8 days period was pixel-averaged across multiple years resulting in 43 layers. The time series of 8-days vegetation and temperature rasters naturally exhibited a very high degree of spatial and temporal correlation. Following the original LKN-methodology we reduced dimensionality and orthogonalized this data by applying Principal Component (PC) decomposition to the original time series. We found that the first 4 and 8 principal components retained 90% and 95% of the original information respectively. A composite image created from the first four components showed a high spatial separation and a large signal to noise ratio (Figure 2B).

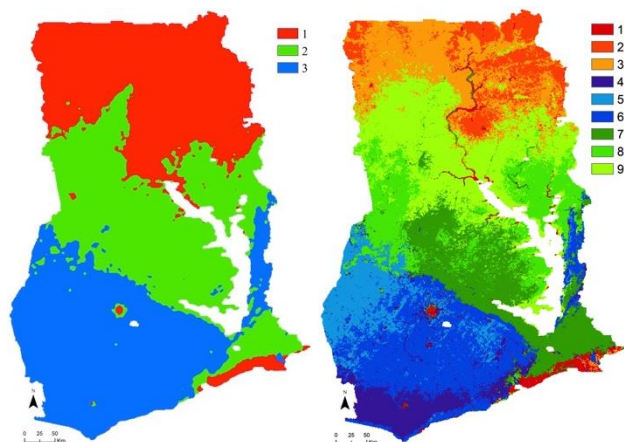
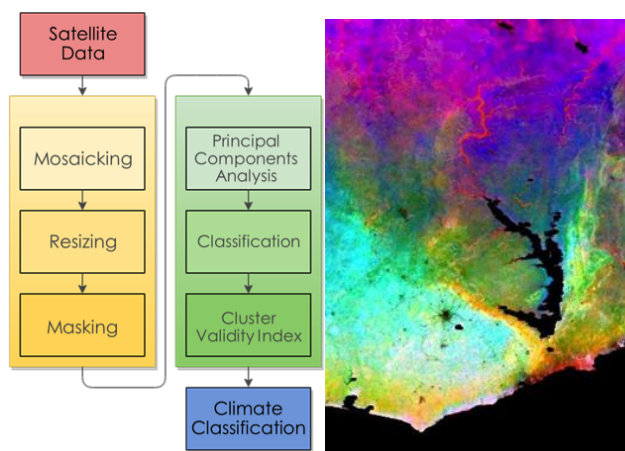


Figure 1: Climate classification workflow. A) flowchart of the LKN method. B) first four components of the PCA. C) 3 major zones and D) 9 minor zones determined by CVI analysis.

The original methodology employs cluster analysis to define regions with similar climate. It aims to assign a finite set of labels (also known as categories or classes) to a very large number of multidimensional objects (pixels, representing a defined area on the ground in our case) based on their similarity followed by determination of the cluster validity. We perform classification using several parameter sets and algorithms with a varying number of principal components, regions, and applied distance measures resulting in approximately 200 classifications. We assessed the quality of clustering using cluster validity index (CVI) analysis. K-Means using Euclidean distance produced the best clustering, with 3 major zones (Figure 2C) and 9 minor zones (Figure 2D).

### 3.2 Associations between disease and RS-based parameters

Spatiotemporal patterns of schistosomiasis disease rates were explored using descriptive statistics, Spearman's correlation, and a multivariate generalized linear mixed effects regression model. The model was designed to control for temporal variations, such as trend and seasonality with two harmonic terms, and spatial variations by including district-level clustering effects. The model was repeated for major and minor climatic zones and formulated as follows:

$$Y_{tj} = \beta_0 + \beta_1 t + \beta_2 S + \beta_{3k} RS_{tj} + \beta_{3j} RS_t + \epsilon_{tj}. \quad (1)$$

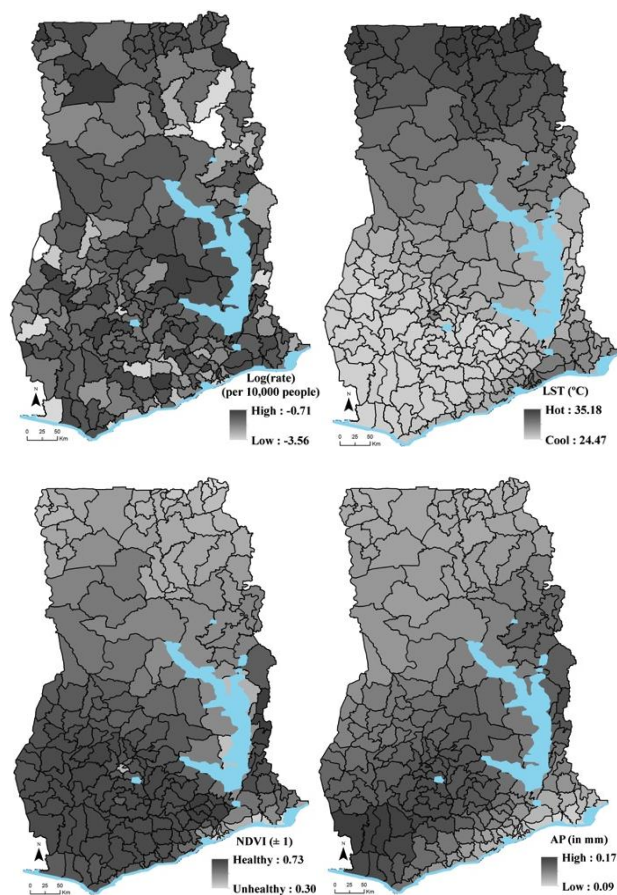


Figure 2: Spatial distribution of the 8 year averages of Loge(rate), LST, NDVI, and AP aggregated from 2008-2015 and extracted by the 216 districts of Ghana. Districts with no counts for the entire time period are shown in white, and data values were stretched from high to low.

where  $Y_{ij}$  is the natural log transformed monthly rate per 10,000 people at  $t$ -month and  $j$ -district;  $t$  is the time in months (96 months from 2008 – 2015);  $\beta_L$  and  $\beta_M$  represent the regression coefficients for fixed effects of seasonality,  $S$ , and remotely sensed variables,  $RS$ , respectively:

$$S = \beta_2 \sin(2\pi\omega t) + \beta_3 \cos(2\pi\omega t) + \beta_4 \sin(4\pi\omega t) + \beta_5 \cos(4\pi\omega t)$$

$$RS = \beta_6 x_{1tj} + \beta_7 x_{2tj} + \beta_8 x_{3tj} \quad (2)$$

$$(3)$$

$b_p$  represents the coefficients of the random effects for the remote sensing parameters for each district. Two harmonic terms account for potential multiple peaks, as the north of Ghana is characterized by one wet-dry season, while the south by two seasons. The models were fitted by the restricted maximum likelihood (REML) using the glmer function of the R package [lme4] (version 3.2.2).

Models results were presented as relative risk (RR) estimates with their 95% confidence intervals (95%CI). Model fit was assessed by superimposing the fitted model on a needle plot representation of the outcome's time series.

#### 4. RESULTS

The general summaries of outcome and predictor variables indicate spatial heterogeneity by climate zone (see Table 3). As evident from the spatial distributions of the predictor variables (Figure 2) there are clear climatic divisions in the long term averages of RS parameters. These divisions were well captured by the LKN regions (Figure 1), which provide a more accurate spatial unit of analysis in assessing the relationship between climate and schistosomiasis than administrative boundaries.

LKN regions affected the long term averages of the outcome and predictor variables. When stratified into the 9 minor regions, the rates ranged from 0.18 to 1.09 per 10,000 people (mean  $\pm$  stdev. =  $0.45 \pm 1.61$ ), LST ranged from 25.64 to 34.03 degrees Celsius ( $28.18 \pm 4.36$ ), NDVI ranged from 0.39 to 0.68 in health of vegetation between  $\pm 1$  ( $0.58 \pm 0.17$ ), and AP ranged from 0.10 to 0.14 in mm/month of rainfall ( $0.13 \pm 0.09$ ). The long term average of reporting rate was very low, and could reflect underreporting.

All values of LST fall within the ideal range of temperature values suitable for the schistosoma parasite and snail species (Gryseels et al., 2006). Overall, NDVI shows healthy vegetation on average, as would be expected in a tropical environment. The range of rain values was small. The apparent albeit mild, differences in average values per LKN region suggest that studies of this nature require precision in order to flush out substantial spatiotemporal relationships in equatorial regions.

| Zone | Rates           | LST              | NDVI            | AP              |
|------|-----------------|------------------|-----------------|-----------------|
| All  | $0.45 \pm 1.61$ | $28.18 \pm 4.36$ | $0.58 \pm 0.17$ | $0.13 \pm 0.09$ |
| 3:1  | $0.49 \pm 1.48$ | $32.31 \pm 5.18$ | $0.42 \pm 0.15$ | $0.11 \pm 0.10$ |
| 3:2  | $0.50 \pm 2.00$ | $28.49 \pm 3.27$ | $0.57 \pm 0.14$ | $0.12 \pm 0.09$ |
| 3:3  | $0.42 \pm 1.49$ | $25.85 \pm 1.92$ | $0.67 \pm 0.11$ | $0.14 \pm 0.09$ |
| 9:1  | $0.26 \pm 0.71$ | $29.47 \pm 3.17$ | $0.41 \pm 0.11$ | $0.10 \pm 0.07$ |
| 9:2  | $1.09 \pm 2.46$ | $34.03 \pm 5.22$ | $0.39 \pm 0.15$ | $0.11 \pm 0.11$ |
| 9:3  | $0.38 \pm 1.03$ | $33.53 \pm 5.53$ | $0.44 \pm 0.16$ | $0.11 \pm 0.11$ |
| 9:4  | $0.34 \pm 0.81$ | $26.18 \pm 1.89$ | $0.65 \pm 0.09$ | $0.13 \pm 0.09$ |
| 9:5  | $0.18 \pm 2.13$ | $25.64 \pm 2.24$ | $0.67 \pm 0.12$ | $0.13 \pm 0.09$ |
| 9:6  | $0.46 \pm 1.35$ | $25.65 \pm 1.74$ | $0.68 \pm 0.10$ | $0.14 \pm 0.09$ |
| 9:7  | $0.65 \pm 1.87$ | $27.93 \pm 2.69$ | $0.60 \pm 0.12$ | $0.12 \pm 0.08$ |
| 9:8  | $0.33 \pm 2.36$ | $28.29 \pm 3.64$ | $0.56 \pm 0.15$ | $0.13 \pm 0.10$ |
| 9:9  | $0.27 \pm 0.71$ | $30.36 \pm 4.79$ | $0.50 \pm 0.16$ | $0.13 \pm 0.11$ |

Table 3: Descriptive Statistics of outcome and predictors.

Overall, the correlation between the monthly values of rates and the environmental variables was near zero (which is confirmed by a more detailed analysis). There was moderate correlation among the environmental variables (Figure 3), which did not warrant collinearity.

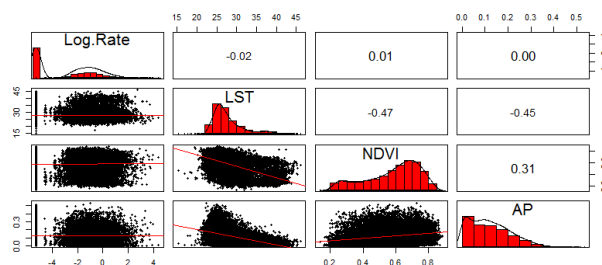


Figure 3: Correlation matrix of outcome and predictors.

The mixed effects models had 41% explained variance overall. Stratification by LKN region brought the  $R^2$  as high as 50% for major zones and as high as 59% for minor zones. Trend and seasonal oscillations contributes a major portion of the variability explained (see Figure 4; 37%). In terms of environmental predictor variables, LST and AP showed no significant risk association with disease rates. NDVI showed increased relative risk values in major zone 3 and minor zone 7 (RR = 1.64; CI95%: 1.01,2.67 and RR = 6.76 CI95%: 1.93,23.62, respectively). AP showed increased relative risk overall (RR = 1.47; CI95%: 1.01,2.15).

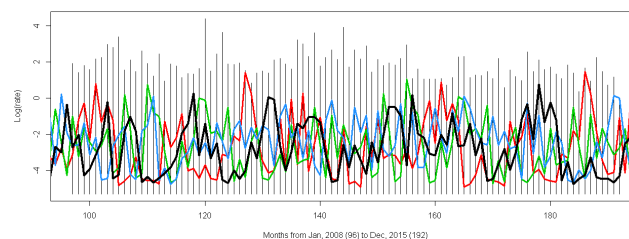


Figure 4: Needle plot of  $\log_e(\text{rate})$  superimposed with fitted models: All (black), Z3:1 (red), Z3:2 (green), and Z3:3 (blue).

|     | LST RR<br>(LCI,UPC) | NDVI RR<br>(LCI,UPC) | AP RR<br>(LCI,UPC)   | Results<br>$R^2$ | n     |
|-----|---------------------|----------------------|----------------------|------------------|-------|
| All | 1.00<br>(0.99,1.01) | 1.32<br>(0.99,1.77)  | 1.47<br>(1.01,2.15)  | 0.41             | 20442 |
| 3:1 | 1.00<br>(0.98,1.02) | 0.57<br>(0.27,1.19)  | 0.56<br>(0.26,1.20)  | 0.50             | 5622  |
| 3:2 | 0.99<br>(0.96,1.03) | 0.96<br>(0.36,2.55)  | 2.29<br>(0.87,6.01)  | 0.39             | 4275  |
| 3:3 | 1.00<br>(0.97,1.03) | 1.64<br>(1.01,2.67)  | 1.09<br>(0.63,1.87)  | 0.37             | 10545 |
| 9:1 | 0.98<br>(0.93,1.04) | 0.56<br>(0.08,4.12)  | 0.86<br>(0.18,3.99)  | 0.42             | 1567  |
| 9:2 | 1.01<br>(0.97,1.06) | 1.46<br>(0.10,20.50) | 0.47<br>(0.06,3.48)  | 0.59             | 1407  |
| 9:3 | 0.96<br>(0.93,1.00) | 0.32<br>(0.04,2.97)  | 0.36<br>(0.06,2.09)  | 0.39             | 1985  |
| 9:4 | 0.97<br>(0.90,1.04) | 0.65<br>(0.15,2.93)  | 0.79<br>(0.20,3.12)  | 0.29             | 1805  |
| 9:5 | 1.00<br>(0.95,1.05) | 2.47<br>(0.96,6.38)  | 1.21<br>(0.32,4.54)  | 0.28             | 1805  |
| 9:6 | 1.00<br>(0.97,1.05) | 1.61<br>(0.79,3.28)  | 1.42<br>(0.69,2.94)  | 0.32             | 5795  |
| 9:7 | 1.00<br>(0.96,1.04) | 6.76<br>(1.93,23.62) | 1.76<br>(0.50,6.24)  | 0.40             | 3418  |
| 9:8 | 0.98<br>(0.93,1.04) | 0.35<br>(0.07,1.63)  | 0.67<br>(0.13,3.51)  | 0.28             | 1425  |
| 9:9 | 0.99<br>(0.93,1.06) | 0.95<br>(0.10,8.76)  | 4.35<br>(0.66,28.83) | 0.47             | 1235  |

Table 4. Results of mixed effects models per LKN regions.

## 5. DISCUSSION

The proposed mixed effects model explained ~ 41% of spatio-temporal variability of the health outcome. When stratified by LKN region the  $R^2$  ranged from 37-50% for the 3 major regions, and 28-59% for the 9 minor regions. RR associated with the trend variable in the models reflects the decline in the reported rates of schistosomiasis over the study period 2008-2015. While the exact nature of this will require further investigation, it is likely that mass drug administration (MDA) campaigns with the deworming drug, praziquantel, are a potential cause. MDA campaigns began in 2008 and have continued through 2015 to present (WHO, 2010; Merck, 2016). Also important to note is that according to the World Bank, Africa currently has one of the highest population growth rates in the world, necessitating the use of accurate population adjustments and growth rates per individual years (Linard et al., 2012). No similar decreasing trends were noted in the predictor variables, which alternatively may show effects of climate change.

Exploratory analyses showed noticeable seasonality in the environmental predictor variables. The documented single wet/dry cycle of northern Ghana was apparent in the major LKN region 1, while the double wet/dry cycles common to the south were seen in the major LKN regions 2 and 3. In addition to showing agreement with known hydrological cycles, the major LKN regions aligned well with the Köppen-Geiger climate classification regions. Thus, remote sensing feeds can effectively be used to capture the seasonal effects and climate divisions. Assessing the nature of schistosomiasis count seasonality needs further assessment; however model fit improved substantially with the incorporation of seasonal harmonics.

A long lag between infection and clinical symptoms may contribute to underreporting (Danso-Appiah et al., 2004; de Vlas et al. 2014), and so obscure the documentation of seasonality. Ghana's equatorial location makes variations in seasonal environmental variables subtle, and so makes the precision of the data in this type of analysis especially important. Of the 600 partitions run on 200 classifications, most agreed that there were only 3 major and 7-10 minor clusters. For these reasons it is difficult to link schistosomiasis rates with purely environmental variables. However the 41% explained variance is promising.

To overcome these challenges, this study has utilized 8 years of overlapping remote sensing data feeds and national health data. Other studies approaching this length of temporal data have relied upon retrospective or prospective health surveys (Walz et al, 2015a). Studies that have taken place on similar or greater spatial scales have relied upon aggregated retrospective health surveys (Shur et al., 2011; Shur et al., 2013). While these data sources are highly accurate for the time and place in which they were collected, extrapolation of these results have increased uncertainty.

Access to the national health data of schistosomiasis in Ghana was granted based on recent research on schistosomiasis and water in Ghana (Kosinski et al. 2011; Kosinaki et al., 2012; Kulinkina et al., 2016) The last systematic collection of health survey data in Ghana at the national scale and subsequently used to direct control efforts was in 2008 (Soares-Magalhaes et al., 2011). This study used detailed health data collected in 77 schools to direct post study drug distribution by interpolating predicted risk for the rest of the country. A purely

environmental approach paired with access to national health data for the reportable disease, schistosomiasis, has the benefit of being a more cost effective spatiotemporal tool.

Increasing spatiotemporal resolutions of environmental and outcome variables would be of benefit to improving model fit (Utzinger et al., 2009). However the MODIS sensor has a spatial resolution limit of 250m unsharpened, and a temporal resolution limit of 4 day adapted composites before alternative satellite sensors would need consideration. ASTER or Landsat would be considered for future publically available satellite data as they provide continuous, high spatiotemporal resolution. However the computational price of this will need careful consideration.

By running the LKN regionalization algorithm hierarchically on each of the major zones separately we can hone in on subtler climatic differences on scales closer to our district level rates. Alternatively the spatial resolution of the health outcome is available at the rural health unit, provided that the georeferencing of the clinics is undertaken. Bringing the environmental and health data to a comparable scale is one source of improved model fit. Additional improvements would come from increasing the overall spatiotemporal resolution of all variables; adding additional environmental variables, use of fourier transform analysis to guide seasonality analysis, and incorporating adjustments such as collinearity, autocorrelation, over dispersion, and lag. By exploring the potential of purely environmental data to predict schistosomiasis, we are seeing if there might be a more sustainable, cost effective and rapid method for disease surveillance.

## ACKNOWLEDGEMENTS

We would like to thank members of Ghana Health Service for providing access to monthly health data from the national surveillance system.

## REFERENCES

- Bavia, M. E., Malone, J. B., Hale, L., Dantas, a, Marroni, L., & Reis, R. (2001). Use of thermal and vegetation index data from earth observing satellites to evaluate the risk of schistosomiasis in Bahia, Brazil. *Acta Tropica*, 79(1), 79–85. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11378144>
- Chitsulo, L., Engels, D., Montresor, a, & Savioli, L. (2000). The global status of schistosomiasis and its control. *Acta Tropica*, 77(1), 41–51. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10996119>
- Danso-Appiah, A., De Vlas, S. J., Bosompem, K. M. and Habbema, J. D. F. (2004), Determinants of health-seeking behaviour for schistosomiasis-related symptoms in the context of integrating schistosomiasis control within the regular health services in Ghana. *Tropical Medicine & International Health*, 9: 784–794. doi: 10.1111/j.1365-3156.2004.01267.x
- De Vlas SJ, Danso-Appiah A, van der Werf MJ, Bosompem KM, Habbema JDF: Quantitative evaluation of integrated schistosomiasis control: the example of passive case finding in Ghana. *Trop Med Int Health* 2004, 9(6):A16-A21.
- Ghana Statistical Service. 2010 Population and housing Census: National analytical report. 2013; [http://www.statsghana.gov.gh/pop\\_stats.html](http://www.statsghana.gov.gh/pop_stats.html).

- Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 2012, 7(2): e31743.
- Gray, D. J., McManus, D. P., Li, Y., Williams, G. M., Bergquist, R., & Ross, A. G. (2010). Schistosomiasis elimination: Lessons from the past guide the future. *The Lancet Infectious Diseases*, 10(10), 733-736.
- Gryseels, B., Polman, K., Clerinx, J., & Kestens, L. (2006). Human schistosomiasis. *The Lancet*, 368(9541), 1106-1118.
- Grosse, S. (1993). *SCHISTOSOMIASIS AND WATER RESOURCES DEVELOPMENT: A RE-EVALUATION OF AN IMPORTANT ENVIRONMENT-HEALTH LINKAGE* (p. 43). Ann Arbor, Michigan. Retrieved from <http://ageconsearch.umn.edu/bitstream/11881/1/twp2.pdf>
- Hürlimann E, Schur N, Boutsika K, Stensgaard A-S, Laserna de Himpsl M, Ziegelbauer K, et al. (2011) Toward an Open-Access Global Database for Mapping, Control, and Surveillance of Neglected Tropical Diseases. *PLoS Negl Trop Dis* 5(12): e1404. doi:10.1371/journal.pntd.0001404
- Kosinski KC, Bosompem KM, Stadecker MJ, Wagner AD, Plummer J, Durant JL, Gute DM: Diagnostic accuracy of urine filtration and dipstick tests for *Schistosoma haematobium* infection in a lightly-infected population of Ghanaian schoolchildren. *Acta Trop* 2011, 118(2):123-127.
- Kosinski, K. C., Adjei, M. N., Bosompem, K. M., Crocker, J. J., Durant, J. L., Osabutey, D., Plummer, J.D., Stadecker, M.J., Wagner, A., Woodin, M., Gute, D. M. (2012). Effective Control of *Schistosoma haematobium* Infection in a Ghanaian Community following Installation of a Water Recreation Area. *PLOS Neglected Tropical Diseases*, 6(7). doi:10.1371/journal.pntd.0001709
- Kosinski, Karen Claire, Alexandra V. Kulinkina, Akua Frimpomaa Atakora Abrah, Michael N. Adjei, Kara Marie Breen, Hafsa Myedah Chaudhry, Paul E. Nevin, Suzanne H. Warner, and Shalini Ahuja Tendulkar. "A Mixed-methods Approach to Understanding Water Use and Water Infrastructure in a Schistosomiasis-endemic Community: Case Study of Asamama, Ghana." *BMC Public Health* 16.1 (2016).
- Kottek, Markus, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. "World Map of the Köppen-Geiger Climate Classification Updated." *Meteorologische Zeitschrift Meteorol. Z.* 15.3 (2006): 259-63. Retrieved from <http://www.srh.noaa.gov/jetstream/global/images/Koppen-Geiger.pdf>
- Kulinkina, Alexandra V., Karen C. Kosinski, Alexander Liss, Michael N. Adjei, Gilbert A. Ayamgah, Patrick Webb, David M. Gute, Jeanine D. Plummer, and Elena N. Naumova. "Piped Water Consumption in Ghana: A Case Study of Temporal and Spatial Patterns of Clean Water Demand Relative to Alternative Water Sources in Rural Small Towns." *Science of The Total Environment* 559 (2016): 291-30.
- Liss, A., Koch, M., & Naumova, E. N. (2014). Redefining climate regions in the United States of America using satellite remote sensing and machine learning for public health applications. *Geospat Health Geospatial Health*, 8(3), 467.
- Martins-Melo, Francisco Rogerlândio, Marta Cristhiany Cunha Pinheiro, Alberto Novaes Ramos, Carlos Henrique Alencar, Fernando Schemelzer De Moraes Bezerra, and Jorg Heukelbach. "Spatiotemporal Patterns of Schistosomiasis-Related Deaths, Brazil, 2000 – 2011." *Emerg. Infect. Dis. Emerging Infectious Diseases* 21.10 (2015): 1820-823. Web.
- Merck KGaA. Fighting Schistosomiasis (accessed 04/2016). <http://www.emdgroup.com/emd/responsibility/society/schistosomiasis/schistosomiasis.html>.
- Schur N, Hürlimann E, Garba A, Traoré MS, Ndir O, Ratard RC, et al. Geostatistical model-based estimates of schistosomiasis prevalence among individuals aged  $\leq 20$  years in West Africa. *PLoS Negl Trop Dis*. 2011;5:e1194.
- Schur N, Hürlimann E, Stensgaard AS, Chimfwembe K, Mushinge G, Simoonga C, et al. Spatially explicit *Schistosoma* infection risk in eastern Africa using Bayesian geostatistical modelling. *Acta Trop*. 2013;128:365–77.
- Soares Magalhaes RJ, Biritwum NK, Gyapong JO, Brooker S, Zhang Y, Blair L, Fenwick A, Clements AC. Mapping helminth co-infection and co-intensity: geostatistical prediction in Ghana. *PLoS Neglected Tropical Diseases* 2011, 5: e1200.
- Simoonga C, Kazembe LN, Kristensen TK, Olsen A, Appleton CC, Mubita P, et al (2008). The epidemiology and small-scale spatial heterogeneity of urinary schistosomiasis in Lusaka province, Zambia. *Geospatial Health*. ;3:57–67.
- Simoonga C, Utzinger J, Brooker S, Vounatsou, P, Appleton CC, Stensgaard AS, Olsen A, Kristensen TK. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology* 2009, 136(13): 1683-1893.
- Steinmann, P., Keiser, J., Bos, R., Tanner, M., & Utzinger, J. (2006). Schistosomiasis and water resources development: Systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases*, 6(7), 411-425.
- Sturrock, R. F., Diaw, O., Talla, I., Niang, M., Piau, J., & Capron, A. (2001). Seasonality in the transmission of schistosomiasis and in populations of its snail intermediate hosts in and around a sugar irrigation scheme at Richard Toll, Senegal. *Parasitology*, 123(07), 77-89. doi:10.1017/S0031182001008125.
- Utzinger, J., Raso, G., Brooker, S., De Savigny, D., Tanner, M., Ornberg, N., ... N'goran, E. K. (2009). Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. *Parasitology*, 136(13), 1859–74. doi:10.1017/S0031182009991600
- Walz Y, Wegmann M, Dech S, Raso G, Utzinger J. Risk profiling of schistosomiasis using remote sensing: approaches, challenges and outlook. *Parasites & Vectors* 2015, 8: 163 (a).
- Walz Y, Wegmann M, Leutner B, Dech S, Vountsou, P, N'Goran EK, Raso G, Utzinger J. Use of an ecologically relevant modelling approach to improve remote sensing-based schistosomiasis risk profiling. *Geospatial Health* 2015, 10: 398 (b).

Walz Y, Wegmann M, Dech S, Vounatsou P, Poda JN, N'Goran EK, Utzinger J, Raso G. Modeling and validation of environmental suitability for schistosomiasis transmission using remote sensing. *PLoS Neglected Tropical Diseases* 2015, 9(11): e0004217 (c).

WHO. Ghana Country Profile 2010 (accessed 04/2016). [http://www.who.int/neglected\\_diseases/preventive\\_chemotherapy/databank/CP\\_Ghana.pdf?ua=1](http://www.who.int/neglected_diseases/preventive_chemotherapy/databank/CP_Ghana.pdf?ua=1).