

Mapping Minimum and Maximum Standard Deviation 3-Dimensional Slope Coefficients for Geo-spectrotemporally Iteratively Quantitatively Interpolating an End-member Proxy Signature of Cyanobacteria (Blue-green Algae) for Eco-cartographically Delineating Cholera Risk in a Riverine Tributary Ecosystem in Ecuador

Sarita Panchang*, Samuel Alao, Jacob GB and Ricardo Izurieta

Department of Global Health, College of Public Health, University of South Florida, Tampa, FL 33613, USA

Abstract

Cholera, caused by the bacteria *Vibrio cholerae*, hit Latin America after nearly a century in 1991. Ecuador was one of the most highly impacted regions by morbidity. While a number of infrastructural and socially relevant factors have been cited in cholera transmission, a growing body of work also points to the relevance of ecological factors, namely the presence and cycle of copepods and blue-green algae which may harbor the bacteria at certain times. The goal of this study was to demonstrate the utility of an interpolation technique which could be used alongside cholera case rate data to determine a predictive environmental signature for cholera risk areas in the future.

Keywords: *Vibrio cholera*; Cyanobacteria; Ecosystem; Mapping

Introduction

Cholera, transmitted by the bacteria *Vibrio cholerae*, is a severe disease that has been documented for centuries. Cholera may produce a range of symptoms involving asymptomatic infection to severe diarrhea and dehydration due to enterotoxins produced by the bacteria which interfere with absorption of key electrolytes. While there is no medicinal cure for the disease, rehydration therapy is a key treatment [1]. Because of the bacteria's persistence in seafood and warm or brackish waters and due to its fecal-oral transmission [2], cholera serves as an indicator of key public health standards such as water and sanitation quality, access to primary health care, and socioeconomic disadvantage. In 2011 alone, 58 countries around the world reported a total of 589854 cases of cholera, an increase from the previous year thought to be associated with the outbreak introduced into Haiti in 2010 following an earthquake [3]. Prior to this most recent epidemic, Latin America saw the first cholera outbreak in nearly a century when the disease was noted in a coastal area of Peru in 1991 [4] and proceeded to spread its way through almost every other country in Latin America and the Caribbean [4]. Ecuador was the second country to report a cholera outbreak, mere weeks after it appeared in Peru [1]. Between 1991 and 1994 alone Ecuador was reported to have 86808 cases [2], and the second-highest morbidity rate of the affected region between 1991 and 1996 [4]. Though case definitions did vary by country [4], rates are generally thought to be under-reported [2]. Analyses on the outbreak in Ecuador have pointed to a number of social mechanisms that may have influenced cholera transmission, including lack of wastewater treatment, illegal water connections, and types of water storage containers used. Some groups, such as males above fifteen years, appeared to be at higher risk because they were more likely to work outside of the home and be exposed, especially through consumption of street food [1].

Spatial analysis proves particularly useful in the analysis of disease spread by accounting for broad ranging factors such as geographic areas of risk, social networks, and weather patterns. From data on a 2008-2009 cholera outbreak in Harare, Zimbabwe, a 100 meter increase in elevation corresponded with a 30% reduction in the attack rate ratio [5]. A study in Matlab, Bangladesh also documented that the

three strongest risk factors during two separate cholera epidemics from the 1980's and 1990's were proximity to surface water, high population density, and poor education level [6]. In Esmeraldas-one of the most severely affected provinces of Ecuador during the initial years of the outbreak [7], density and community connectedness were associated with increased incidence of diarrheal disease, but surprisingly factors such as proximity to roads were associated with more community dispersal and thus lower risk of cholera spread [8]. Another nationwide spatial analysis using smoothed incidence rates indicates that the cholera epidemic in Ecuador had two major peaks in early 1991 and late 1992. Between 1991 and 1996, the province of Guayas contained about one third of all cases. Two thirds occurred in the coastal provinces more generally, with the provinces of Imbabura and El Oro having some of the highest average weekly incidence rates. One epicenter constituted the southern provinces – El Oro (the origin), Guayas, Canar, Loja, and Los Rios. The second epicenter was Esmeraldas and Imbabura. This is an interesting finding considering that some of these provinces, such as Imbabura, are located near the highlands and are located in a higher altitude, and exhibits the utility of using geographic features to pinpoint potential cholera outbreaks in diverse ecological conditions which may favor the bacteria's growth [7].

It is known that seasonal patterns of cholera abundance can be attributed to environmental and climactic cycles that also influence copepods which frequent marine and estuarine systems and have

*Corresponding author: Sarita Panchang, Department of Global Health, College of Public Health, University of South Florida, Tampa, FL 33613, USA, Tel: +8139742011; E-mail: panchang@health.usf.edu

Received November 04, 2016; Accepted November 17, 2016; Published November 21, 2016

Citation: Panchang S, Alao S, Jacob GB, Izurieta R (2016) Mapping Minimum and Maximum Standard Deviation 3-Dimensional Slope Coefficients for Geo-spectrotemporally Iteratively Quantitatively Interpolating an End-member Proxy Signature of Cyanobacteria (Blue-green Algae) for Eco-cartographically Delineating Cholera Risk in a Riverine Tributary Ecosystem in Ecuador. J Remote Sensing & GIS 5: 180. doi: [10.4180/2469-4134.1000180](https://doi.org/10.4180/2469-4134.1000180)

Copyright: © 2016 Panchang S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

been found to host the bacteria [9]. In particular, certain types of cyanobacteria – also known as blue-green algae – may play a key role in maintaining endemicity of cholera [10-12]. A remote sensing technique was used by deMagny et al. which uncovered a relationship between higher chlorophyll a levels in coastal areas near Kolkata and Matlab, Bangladesh and algal blooms as well as copepod abundance, predicting recent cholera outbreaks. Since coastal regions of Ecuador were especially affected during the epidemic of the 1990's throughout South America, predictive modeling of cholera risk through identification of key geospatial features would be an essential tool in early detection and intervention for future epidemics. This study demonstrates the utility of an interpolation technique to develop an endmember signature for blue-green algae based on known cholera case rates, a novel usage of remote sensing in water-borne disease.

Methodology

Ecuador is generally divided into three major ecological and climatic zones: the coast, the highlands, and the eastern Amazon [7]. However, even within these categories – sometimes in the same latitude – there is a wide range of weather and climate conditions, and topographical variation (Figure 1). Along the coast, rainfall can vary from 4000 mm per year in the north to 100 mm in the south (Third Millennium Alliance 2015). In 2015 the population of the country was estimated at 16,144,363, though density varies considerably as well (World Bank 2016). Mean elevation data obtained from GMTED indicates a minimum of -4 and a maximum of 6202.

For this study, imagery was obtained from Google Earth and Landsat 7 and 8, obtained for optimal cloud cover and dates covering key months of the year just prior to the rainy seasons when algal blooms are most likely to occur. The resulting Landsat™ images were obtained in November and May 2015. The southern province of Guayas was chosen to apply this interpolation technique (Figure 1) due to its coastal location and high abundance of mangrove and estuarine habitats. Prior literature (see above) as well as chloropleth maps (Figure 3) of cumulative incidence of cholera from 1991 and 1992 (Ecuadorian Ministry of Health, Ricardo Izurieta, personal communication) also indicate that Guayas was one of the most severely affected provinces of the country.

Environmental parameters

Cholera incidence rates per thousand were cartographically overlaid onto the study site in ArcGIS region in order to determine actual rates at the eco-epidemiological riverine study site (Figure 2). Multiple georeferenced predictors were then examined extensively using georeferenced and altitude data. The criteria involved the centographic measures of spatial mean and distance between a sampled, canopy, shaded, bloom of blue-green algae (i.e., capture point) and the distance from the site to the nearest human habitation village sample and to other habitats. The data bloom together with a battery of categorical attribute measures which were expanded into multiple spectrally-dependent, explanatorial, endmember, covariate, coefficient estimates.

The bloom distances were then measured as Euclidean distances in the projection units of the raster which were computed within the digitized grid cell matrix. The Euclidean distance output raster contained the measured distances. The Euclidean Distance functions provided information according to Euclidean or, straight-line, distance between the georeferenced bloom and from bloom to human habitation (i.e., geometric distances in the multidimensional space). The Euclidean

distances were computed as: distance () $21/2 \times x, y \{ () \} = -\sum_i x_i y_i$. Every cell in the Euclidean allocation output raster was assigned the value of the source to which it was closest. The nearest source was determined by the Euclidean Distance function in ArcGIS. This function assigned space between the georeferenced, canopy, shaded, geo-classified land use land cover (LULC) bloom of blue-green algae. The Euclidean

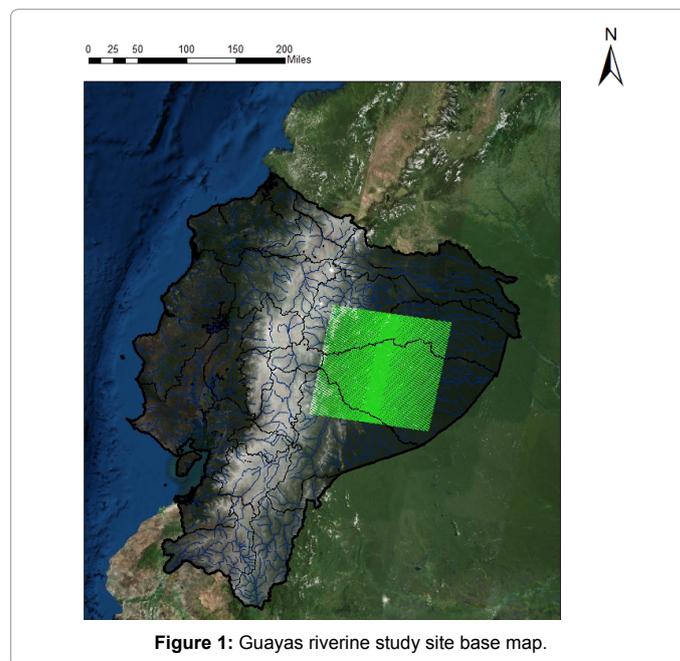


Figure 1: Guayas riverine study site base map.

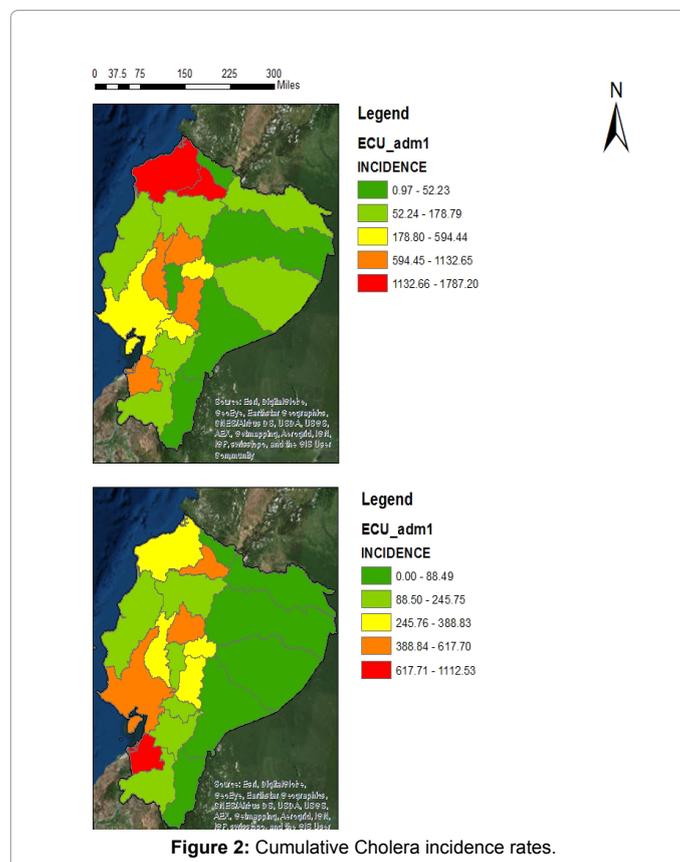


Figure 2: Cumulative Cholera incidence rates.

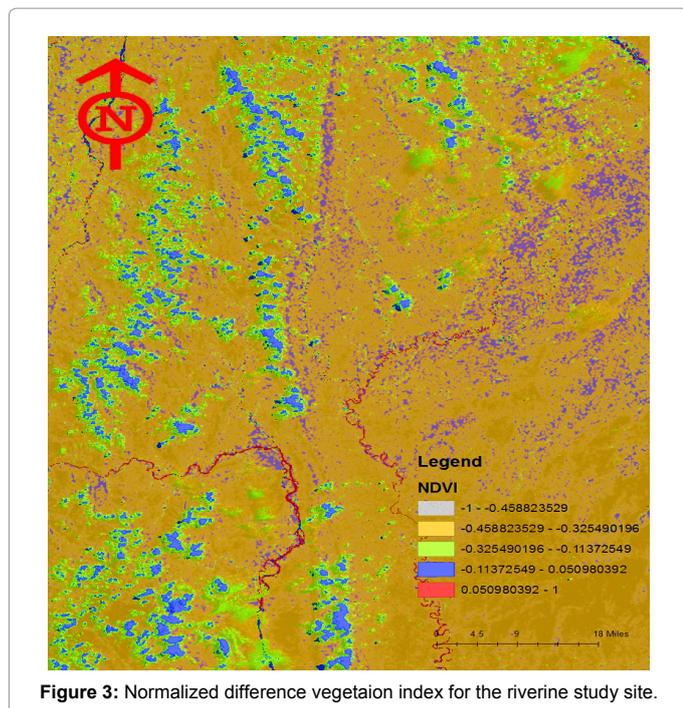


Figure 3: Normalized difference vegetation index for the riverine study site.

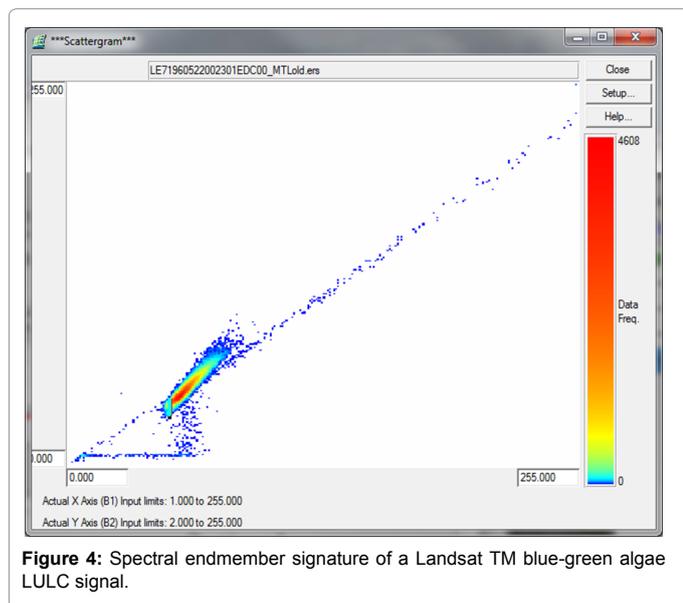


Figure 4: Spectral endmember signature of a Landsat TM blue-green algae LULC signal.

direction output raster contained the azimuth direction from each grid cell centroid to the nearest source. The Euclidean Allocation function identified the nearest human habitation center closest to each grid cell. The distance between sampled blue-green algae bloom LULCs and human habitation areas were then categorized into specific Euclidean-distance based classes (e.g., 1: 0-5 km, 2: 5-10 km, and so on).

Ecohydrological and vegetation models

The latest version of PCI Geomatics Orthoengine[®] software was then used to generate a digital elevation model (DEM) from the spatiotemporal-sampled, georeferenced, canopy, shaded, blue-green algae LULC. PCI software supports automatic overlay of vector arthropod-related habitat DGPS collections, geometric modeling

using Toutins rigorous model, Rationale Polynomial Coefficients (RPC) models, automatic DEM generation, orthorectification and automatic mosaicking [1]. Orthoengine also offers an industry-leading variety of control sources, including manual entry, geocoded imagery, geocoded vectors, chip database, digitizing tablet or a text file (www.pigeomatics.com).

We employed differentially corrected ground control points (DGCP's) from a CSI max receiver which according to Jacob et al. has a positional accuracy of 0.178 m of the sampled blue-green algae to orient the images to our mapping coordinate system. Our RPC method used an empirical/statistical model which approximated a 3-D physical sensor model of Landsat TM visible and near infra-red (NIR). Since bias or error may have still existed in the RPCs, the results were post-processed with a polynomial adjustment product using the DGCPs seasonal-sampled, canopy, shaded, blue-green algae LULC. The original RPC parameters were refined with linear equations and the DGPS data. We used:

$$\Delta P = A_0 + A_S + \text{Sample} + A_L \cdot \text{Line} + A_{SL} \cdot \text{Sample} \cdot \text{Line} + \dots$$

$$\Delta R = B_0 + B_S + \text{Sample} + B_L \cdot \text{Line} + B_{SL} \cdot \text{Sample} \cdot \text{Line} + \dots$$

where A_0 , A_S , A_L , A_{SL} ... and B_0 , B_S , B_L and B_{SL} ... were the image adjustment parameters. Since bias or error may have still existed after applying the RPCs, the results were post-processed. The original RPC parameters were refined with linear equations and the DGCPs. The ΔP and ΔR were the adjustable functions expressing the difference between the measured and the nominal line and the sample coordinates. For most images, zero-order polynomial adjustment (e.g., A_0 and B_0) are required (www.digitalglobe.com). The OrthoEngineTM software supported both zero and first order RPC polynomial adjustments.

The different modules in Spatial Analyst extension of ArcGIS 10.2[®] was used along with spatial modeller tools from ERDAS Imagine 9.3[®] to perform VI calculations. NDVI was calculated using radiance, surface reflectance (p), and apparent reflectance values in the red (0.63 to 0.69 μm) and NIR (0.76 to 0.90 μm) spectral bands. The ratio of reflected radiance from the red and NIR bands were used to normalize illumination and topographic variation and to form the NDVI, which was then used as an indicator of the amount and vigour of vegetation in the riverine epidemiological study site. Initially, a sensitivity analysis was conducted prior to generating NDVI parameters by analyzing the atmospheric and soil-perturbed responses as a continuous function of plant Leaf Area Index (LAI). Leaf area index (LAI) is the total one-sided area of leaf tissue per unit ground surface area [1]. LAI is a key parameter in ecophysiology, especially for scaling up the gas exchange from leaf to canopy level, which normally characterizes the canopy-atmosphere interface, where most of the energy fluxes exchange. LAI can be determined directly by taking a statistically significant sample of foliage from a plant canopy, measuring the leaf area per sample habitat plot and dividing it by the plot land surface area [2]. Analysis of the literature shows that most cross-validations between direct and indirect methods have pointed to a significant underestimation of (LAI) especially in forest stands [7]. The two main causes for canopy, shaded, spectral discrepancy is clumping and contribution of stem and branches however, recent theoretical and technical solutions have been presented as potential improvements to reduce bias or discrepancies [7]. The accuracy, sampling strategy and spatial validity of the (LAI) measurements has to be constantly assessed for quality assurance of both the measurement and for modeling purposes for all calculated (LAI) -dependent ecophysiological and biophysical processes of seasonal, blue-green algae LULC canopies [1]. Canopy geometry is

directly related to LAI. As such, the indicator plays an essential role in theoretical production ecology [8].

The LAI was used to generate photosynthetically, active riverine, blue-green algae LULC canopy, shade-related, explanatory, predictor, covariate, coefficient bloom estimates for quantitating vegetation related LULC parameters at the epidemiological study site. An inverse exponential relation between LAI and light interception was established using $P = m \text{Pax}(1 - \text{ec-LAI})$, where P_{max} designated the maximum primary production and c designated a vegetation specific growth coefficient. Physiological processes such as photosynthesis, transpiration and evapotranspiration are related to LAI [www.esri.com]. Randomly selected grid cell and spectral measurements were assessed to determine plant LAI in ArcMap*.

Estimations of LAI production were conducted by correlation analysis with spectral reflectance ratio and measured blue-green algae LULC values. The best fitting waveband ratio among calculated reflectance and VI's were selected. Percent relative error and vegetation equivalent 'noise' (VEN) were calculated for soil and atmospheric influences in the study site separately and combined using calculated LAI measurements. The NDVI had a relative error of 10 percent and VEN of ± 0.91 LAI.

We then performed Raster modeling in ArcGIS 10.2' which included performing image differencing on the NDVI layers, classifying the layers into different classes and calculating a wetness index employing the Raster Calculator. The difference of the Landsat TM visible and NIR bands was divided by their sum, which formed the functionally equivalent NDVI over the terrestrial surfaces of the riverine epidemiological study site. NDVI spectral variability (i.e., standard deviation of sub-grid pixel estimates) was able to differentiate multiple LULCs. NDVI parameters, was computed directly without any bias or assumptions regarding plant physiognomy, land cover class, soil type, or climatic conditions, within a range from -1.0 to 1.0 using visible and NIR reflectance, (ρ), utilizing and the expression:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}}$$

$$u_{\text{cal}}^2(\text{NDVI}) - \left(\frac{\partial \text{NDVI}}{\partial \rho_{\text{NIR}}} \right)^2 u_{\text{cal}}^2(\rho_{\text{NIR}}) + \left(\frac{\partial \text{NDVI}}{\partial \rho_{\text{red}}} \right)^2 u_{\text{cal}}^2(\rho_{\text{red}}) + 2 \frac{\partial \text{NDVI}}{\partial \rho_{\text{NIR}}} \frac{\partial \text{NDVI}}{\partial \rho_{\text{red}}} \bullet u_{\text{cal}}(\rho_{\text{NIR}} \bullet \rho_{\text{red}})$$

$$\frac{\partial \text{NDVI}}{\partial \rho_{\text{NIR}}} = \frac{2\rho_{\text{red}}}{(\rho_{\text{NIR}} + \rho_{\text{red}})^2}$$

$$\frac{\partial \text{NDVI}}{\partial \rho_{\text{red}}} = \frac{-2\rho_{\text{NIR}}}{(\rho_{\text{NIR}} + \rho_{\text{red}})^2}$$

$$\frac{\partial \text{NDVI}}{\partial \rho_{\text{NIR}}} \frac{\partial \text{NDVI}}{\partial \rho_{\text{red}}} = \frac{-4\rho_{\text{NIR}}\rho_{\text{red}}}{(\rho_{\text{NIR}} + \rho_{\text{red}})^4}$$

To assess the accuracy of the NDVI thematic maps, a simple random sampling method was chosen to ensure sampled selection. To

provide a statistically efficient assessment of accuracy, a conservative sampled size equation was then employed to calculate the sample sizes using an error matrix.

For each mapping region, stratified sampling formulas were applied to estimate spectral error proportions in the ArcGIS vegetation-related LULC maps generated using the estimates of overall and class-specific user's and producer's accuracy. The use of stratified formulas is important for validating field-sampling methods. Accuracy results were computed thereafter through weighting the cell proportions by the proportion of each Landsat™ classified LULC within a given region. Specifically, the overall producer's accuracy was estimated employing the post-stratified formulas. We used post-stratified LULC estimators as the known Landsat™ pixel totals for each land-cover class (N_{i+}) were treated as a stratified random sample of n_{i+} pixels from the N_{i+} pixels in that class. In this research user's accuracy (U_i) were based on the random sampling formulas:

$$\hat{P} = \frac{1}{N} \sum_{k=1}^q \frac{N_{k+}}{n_{k+}} n_{kk}$$

$$\hat{P}_{U_i} = \frac{n_{ii}}{n_{i+}}$$

$$\hat{P}_{A_j} = \frac{(N_{j+} / n_{j+}) n_{jj}}{\sum_{k=1}^q (N_{k+} / n_{k+}) n_{kj}}$$

The producer's accuracy was calculated to determine the probability that a reference sampled Landsat™-classified LULC correctly mapped and measured the errors of omission (1-producer's accuracy). In contrast, the user's accuracy indicated the probability that a sample from the blue-green algae LULC map actually matched the reference data and measured the error of commission (1-user's accuracy). Kappa statistics were then calculated using SAS PROC FREQ. Kappa statistics (κ) are appropriate for testing whether agreement exceeds chance levels for binary and nominal ratings for remotely-sensed models constructed from spatiotemporal-sampled, spectrally-dependent, georeferenced, seasonal, blue-green algae bloom-related, explanatory predictor variables [2]. The equation for κ was:

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

where $\text{Pr}(a)$ was the relative observed agreement among the uncertainty based statistics and $\text{Pr}(e)$ was the hypothetical probability of chance agreement between the DGCP-sampled, georeferenced, observational, blue-green algae LULC data. If the residual estimates were in complete agreement, then $\kappa=1$. If there was no agreement among the classified data other than what would be expected by chance, then $\kappa \leq 0$.

Object-Oriented classification

Once a robust dataset of the remotely-dependent, explanatory, predictor, covariate coefficients was constructed, ENVI[®] spectral tools were used to analyze the LULC data of the canopy, shaded blue-green algae, eco-georeferceable, capture point. We input the data into ENVI technology using the GeoTIFF format. ENVI supports Input File Landsat™ functions including GeoTIFF, NITF and Tile Product (.til) (www.itvis.com/portals/0/pdfs/envi/Getting_Started_with_ENVI.pdf). In ENVI 4.6', a spectrum plot, known as a z-profile, of the pixel

under the cursor was run through all bands of the Landsat™ image. The basic workflow involved importing the data collected in the field from the riverine eco-epidemiological study site into a spectral library. The library was then employed in the endmember collection workflow to perform a supervised classification, based on the sampled riverine blue-green algae LULC spatial data feature attributes. Binary Encoding, Spectral Angular Mapping (SAM) and Spectral Feature Fitting were used to rank and match any unknown spectrum to the materials in the library.

The endmembers of the georeferenced canopy shade-related riverine blue-green algae LULC and its associated attributes were then extracted from ENVI's spectral library. Several spectra corresponding to the different backgrounds in the canopy, shaded, structures, rippled water pixel components had to be included, since multiple scatterings between floating leaves in the habitat, and a bright soil background increased the Landsat™ NIR reflectance. Leaf cells have evolved to scatter (i.e., reflect and transmit) solar radiation in the NIR spectral region [www.esri.com]. After the calibration estimates had been determined the image was converted to match the library. Analogously, the Landsat™ reference endmembers spectra in the library were transformed into the endmembers spectra of the image.

We used a random selection of pixels to assess the satellite classification accuracy. Landsat™ class representative pixels were thereafter selected and compared to a reference training dataset. During the segmentation procedure, image objects were generated based on several adjustable criteria of homogeneity such as colour, shape, and texture. Data pre-processing involved converting the DN to radiance atmospheric correction using FLAASHTM, and co-registration techniques. Image classification was conducted thereafter using the object-oriented approach. FLAASHTM generated a model (.gmd file) that converted the image's DN to at-sensor radiance and computed at-sensor reflectance while normalizing the solar elevation angle.

The equation was as follows:

$$\rho_{\text{BandN}} = \frac{\pi(L_{\text{BandN}} \times \text{GainBandN} + \text{BiasBandN}) \times D^2}{E_{\text{BandN}} \times (\text{COS}((90 - \theta) \times \pi / 180))}$$

Where,

ρ_{BandN} = Reflectance for Band N

L_{BandN} = Digital Number for Band N

D = Normalized Earth-Sun Distance

E_{BandN} = Solar Irradiance for Band N

The reference data in this research was the "ground truthed" data of the selected canopy, shaded, riverine, blue-green algae LULC explanatory, predictor variables. Selected random pixels from the thematic map were then compared to the reference data. The estimated accuracy for the spectrally extracted data was calculated:

$$P\left\{-z_{\frac{\alpha}{2}} < \frac{x - n\theta}{\sqrt{n\theta(1-\theta)}} < z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

where, x=number of correctly identified pixels, n=total number of pixels in the sample, θ =the map accuracy, $(1-\alpha)$ =a confidence limit. The Landsat™ classifier identified the canopy shaded blue-green algae LULC and their associated ripple water larger than 0.4 has with generally satisfactory results (92.1%) with a lower detection limit of 0.1 ha.

The successive projection algorithm (SPA)

We used an SPA to generate endmembers from the blue-green algae LULC and the ripple water spectra. We let $P(i, j)$ denote the spectrum for the spectrally extracted Landsat™ pixel using the image coordinates (i, j) , as the foundation of the unmixing algorithm which in this research was defined by using $\bar{p}_{(i,j)} = \sum_{k=1}^m f_{(i,j)k} \bar{e}_k + \bar{e}_{(i,j)}$ and also $f_{(i,j)k} \geq 0, k = 1, \dots, m, \sum_{k=1}^m f_{(i,j)k} = 1$ where m was the number of endmembers, \bar{e}_k was the k th endmember, $\bar{e}_{(i,j)}$ was the approximation error term (i.e., residual), and $f(i, j)k$, was the fractional abundance for the k th endmember of pixel (i, j) .

We omitted the error term in $\bar{p}_{(i,j)} = \sum_{k=1}^m f_{(i,j)k} \bar{e}_k + \bar{e}_{(i,j)}$ and computed the possible linear mixtures from $\bar{p}_{(i,j)} = \sum_{k=1}^m f_{(i,j)k} \bar{e}_k + \bar{e}_{(i,j)}$ and $f_{(i,j)k} \geq 0, k = 1, \dots, m, \sum_{k=1}^m f_{(i,j)k} = 1$ which formed a simplex C_m defined by m vertices that corresponded to the blue-green algae LULC endmembers, $\bar{e}_1, \bar{e}_2, \dots, \bar{e}_m$. According to Jacob et al. the volume of the simplex C_m can be calculated from the equation $V(C_m) = \frac{1}{(m-1)!} |\det(WW^T)|^{\frac{1}{2}}$ where $W = [\bar{e}_2 - \bar{e}_1, \bar{e}_3 - \bar{e}_1, \dots, \bar{e}_m - \bar{e}_1]$ is the volume of the simplex defined by m endmembers, and where $\det(\cdot)$ denotes the determinant of a matrix representing the operation of an absolute value. Once the endmembers $\bar{e}_1, \bar{e}_2, \dots, \bar{e}_m$ were determined while their abundance was estimated through the least squares method which was equivalent to a projection on the simplex.

We then used the Li-Strahler geometric-optical model based on the assumption that the BRDF would retrieve blue-green algae LULC, riverine canopy structural variables. Initially we generated a bidirectional reflectance distribution function (BRDF). The bidirectional reflectance distribution function is a four-dimensional function that defines how light is reflected at an opaque surface (Jensen). This function took incoming light directions, ω_i , and outgoing directions, ω_o , both defined with respect to the georeferenced algae bloom surface normal n , and returned the ratio of reflected radiance exiting along ω_o to the irradiance incident on the riverine ripple water spectral components from direction ω_i . Note that each direction ω was itself parameterized by azimuth angle ϕ and zenith angle θ , therefore the BRDF as a whole was 4-dimensional. The BRDF we calculated had units sr^{-1} , with steradians (sr) being a unit of solid angle. Therefore the BRDF was defined by: where L was the radiance, E was the irradiance, and θ_i was the angle made between ω_i and the blue-green algae LULC and its associated ripple water surface reflectance emissivities. The BRDF defined how light was reflected and the function took an incoming light direction, ω_i , and outgoing direction, ω_o , which were both defined with respect to the blue-green algae LULC and its within canopy and ripple water surface n , spectral values and returned the ratio of reflected radiance exiting along ω_o to the irradiance incident from direction ω_i .

The inverted Li-Strahler geometric-optical model was then used to retrieve specific spectral uncoalesced blue-green algae LULC predictor covariate coefficients. The reflectance associated with a georeferenced capture point was treated as an area-weighted sum of four fixed reflectance components: sunlit canopy, sunlit background, shaded canopy, and shaded background. In geometric-optical models these four components can be simplified to three: sunlit canopy-C, sunlit background-G and shadow-T [www.esri.com]. The sub-pixel endmember spectral components were derived using G, C, T components' classes which were initially estimated by the Quick Bird image using ENVI, an object-based classification method. For inverting the model, parts of the three spectral components were represented by

(Kg) which was calculated using:

$$K_g = e^{-\pi \cdot M [\sec(\theta_i) + \sec(\theta_v) - O(\theta_i, \theta_v, \phi)]} \quad (2.1)$$

$$\Gamma(\sec \theta_i + \sec \theta_v) \quad (2.2)$$

$$\cos t = \frac{h |\tan \theta_i - \tan \theta_v \cos \phi|}{r (\sec \theta_i + \sec \theta_v)} \quad (2.3)$$

$$M = \frac{-\ln(K_g)}{(\sec \theta_i + \sec \theta_v) (\pi - t + \cos t \sin t)} \quad (2.4)$$

$$CC = 1 - e^{-\pi \cdot M} \quad (2.5)$$

where, i and q were the zenith angles of illumination and viewing, O was the average of the overlap function between illumination and viewing shadows of eco-georeferenceable algae blooms and its associated and ripple water components as projected onto the background. In the geometrical, optical paradigm j was the difference in azimuth angle between illumination and viewing.

Interpolation analyses

Spatially elucidative linear predictors were generated from the blue-green algae LULC and its rippled water components using an Ordinary kriged-based interpolator. The dependent variable was the explanatory spectral emissivity estimates Landsat™ band ratio of the riverine, eco-epidemiological, eco-georeferenceable, capture point rendered from the decomposition of the Landsat™ pixel. We transformed the emissivities to fulfill the diagnostic normality test prior to performing kriging. Interpolation algorithms can generate accurate autoregressive endmember endemic transmission-oriented geopredictive risk maps [www.esri.com].

The unknown value $Z(x_0)$ was interpreted as a random variable located in (x_0) , as well as the values of neighbor's samples $(Z(x_i))$ $i = 1, \dots, N$. The estimator $Z(x_0)$ was also interpreted as a random variable located in the interpolator x_0 , as result of the linear combination of the blue-green algae LULC, spectral endmember predictor variables. In order to deduce the kriging system for the assumptions of the model, the following error committed while estimating $Z(x)$ was estimated employing

$$e(x_0) = \hat{Z}(x_0) - Z(x_0) - [W^T - 1] \cdot [Z(x_1) \dots Z(x_N) Z(x_0)]^T - \sum_{i=1}^N W_i(x_0) \times Z(x_i) - Z(x_0)$$

The quality criteria was then expressed in terms of the mean and variance of the new random variable $e(x_0)$. Since the random function was stationary, [i.e., $E(Z(x_1)) - E(Z(x_0)) = m$], the following constraint was observed:

$$E(e(x_0)) = 0 \Leftrightarrow \sum W_i(x_0) \times E(Z(x_i)) - E(Z(x_0)) = 0 \Leftrightarrow$$

$\Leftrightarrow m \sum w_i(x_0) - m = 0 \Leftrightarrow \sum w_i(x_0) = 1 \Leftrightarrow [W^T - 1] \cdot [Z(x_1) \dots Z(x_N) Z(x_0)]^T = 0$ The variance of the geopredictive endemic transmission-oriented model was then minimized employing $2 \cdot 0 \cdot 0 \in Ex$. Two estimators had $0 \cdot 0 \in = x$, but the dispersion around their mean determines the difference between the quality of blue-green algae LULC estimators [1]. Our model rendered the following

$$\text{Var}(e(x_0)) = \text{Var}([W^T - 1] \cdot [Z(x_1) \dots Z(x_N) Z(x_0)]^T) = [W^T - 1] \cdot \text{Var}([Z(x_1) \dots Z(x_N) Z(x_0)]^T) \cdot [W^T - 1]$$

and $\text{Var}(e(x_0)) = [W^T - 1] \cdot \begin{bmatrix} \text{Var}_{x_1} & \text{Cov}_{x_1 x_0} \\ \text{Cov}_{x_1 x_0}^T & \text{Var}_{x_0} \end{bmatrix} \cdot [W^T - 1]$ where the literals

$\{\text{Var}_{x_1}, \text{Var}_{x_0}, \text{Cov}_{x_1 x_0}\}$ represented

$$\{\text{Var}([Z(x_1) \dots Z(x_N)]^T), \text{Var}(Z(x_0)), \text{Cov}([Z(x_1) \dots Z(x_N)]^T, Z(x_0))\}$$

. We defined $\gamma(h)$ (i.e., variogram) in the endemic endmember transmission-oriented model in the analysis of $Z(x)$. We then wrote an expression for the estimation variance of the canopy, shaded, blue-green algae LULC estimator using the function of the covariance between the samples and the covariances to estimate:

$$\left\{ \begin{aligned} &\text{Var}(e(x_0)) - W^T \cdot \text{Var}_{x_1} \cdot W - \text{Cov}_{x_1 x_0}^T \cdot W - W^T \cdot \text{Cov}_{x_1 x_0} + \text{Var}_{x_0} \\ &\text{Var}(e(x_0)) - \text{Cov}(0) + \sum_i \sum_j w_i w_j \text{Cov}(x_i, x_j) - 2 \sum_i C(x_i, x_0) \end{aligned} \right\}$$

Which minimized $W^T \cdot \text{Var}_{x_1} \cdot W - \text{Cov}_{x_1 x_0}^T \cdot W - W^T \cdot \text{Cov}_{x_1 x_0} + \text{Var}_{x_0}$.

$$\begin{bmatrix} W \\ \mu \end{bmatrix} - \begin{bmatrix} \text{Var}_{x_1} & 1 \\ 1^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \text{Cov}_{x_1 x_0} \\ 1 \end{bmatrix} - \begin{bmatrix} \gamma(x_1, x_1) \dots \gamma(x_1, x_n) & 1 \\ \vdots & \vdots \\ \gamma(x_n, x_1) \dots \gamma(x_n, x_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \\ 1 \end{bmatrix} \text{ Where}$$

the additional parameter μ was a Lagrange multiplier employed in the minimization of the kriging error $2 \cdot 0 \cdot \sigma$ to honor the unbiasedness condition. Finally, the estimation and error by Ordinary kriging was given by:

$$\hat{Z}(x_0) - W^T \cdot [Z(x_1) \dots Z(x_N)]^T : (\hat{Z}(x_0) - Z(x_0)) - \hat{W}^T \cdot [\gamma(x_1, x_0) \dots \gamma(x_n, x_0)]^T$$

. Additionally, we tested our model residual forecast with that of current unmixing algorithms commonly applied for canopy-oriented blue-green algae LULC endmember signature generation. This unmixing algorithm was based on the following linear mixing model:

$$R_k = \sum_{i=1}^n a_i \cdot E_{i,k} + \epsilon_k \quad \text{RMSE} = \sqrt{\left(\sum_{k=1}^m \epsilon_k^2 \right)^{-m}}$$

R_k : Reflectance of source at wavelength k

$E_{k,i}$: Reflectance of endmember i at wavelength k

a_i : Abundance of endmember i

ϵ_k : Error at wavelength k

RMSE: Root Mean square error of the ϵ_k

N : Number of endmembers

M : Number of wavelengths in the discrete spectrum

This model assumed that spectrum is a linear superposition of the blue-green algae LULC shaded canopy-related endmembers. The abundances were computed using two standards the algorithm assumed that wavelengths were given in nanometers (i.e., BEAM's default wavelength unit). The spectral unmixing tool was invoked from VISAT's tool menu by selecting the Unmix command and by using the command line tool gpt (Graph Processing Tool) found in BEAM's bin directory in batch mode (Figures 4 and 5). The differential form of the equation for the standard spectral unmixing, diagnostic, eco-georeferenceable, forecast, vulnerability, sub-pixel, model was:

$$\frac{1}{C} \frac{\partial}{\partial t} I_v + \hat{\Omega} \cdot \nabla I_v + (k_{v,s} + k_{v,a}) I_v = j_v + \frac{1}{4\pi C} k_{v,s} \int_{\Omega} I_v d\Omega$$

where j_v was the georeferenced, canopy, shaded, riverine, blue-green algae LULC emission coefficient, $k_{v,s}$ was the scattering cross section, and $k_{v,a}$ was the absorption cross section. We noticed that in our model the differences were essentially due to the various forms for the canopy-related emission and absorption coefficients. The emission coefficient and absorption coefficient are functions of temperature and density are related by: $\frac{j_k}{a_k} = B_k(T)$ where B_k is the black body spectral radiance at temperature T [www.esri.com]. The solution to the equation of our predictive radiative transfer was then:

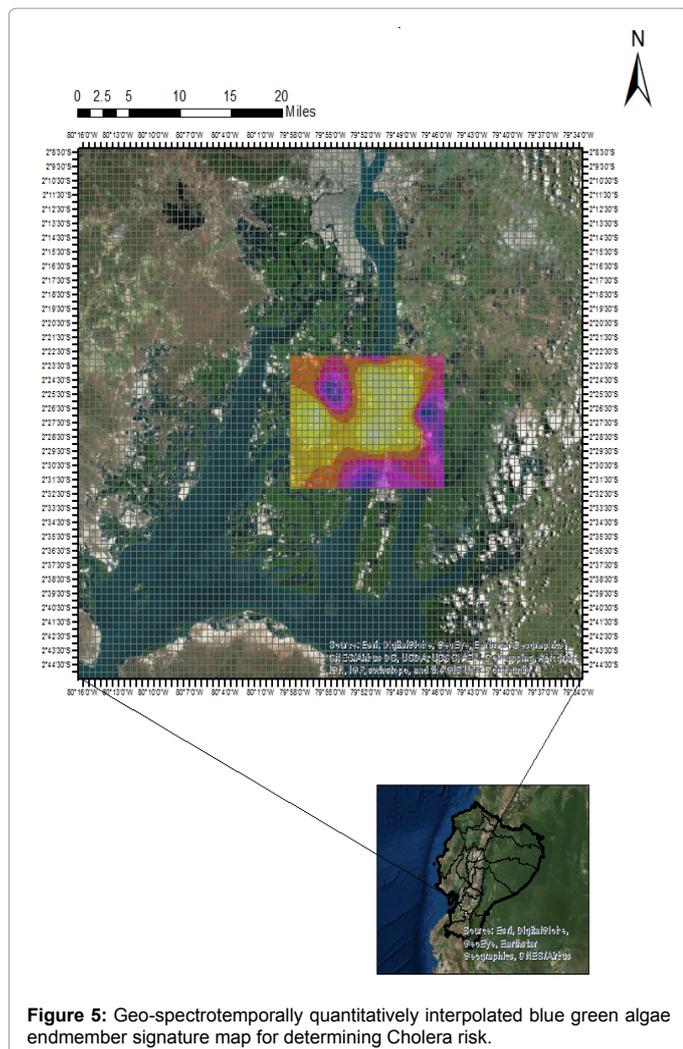


Figure 5: Geo-spectrotemporally quantitatively interpolated blue green algae endmember signature map for determining Cholera risk.

$$I_v(s) = I_v(s_0) e^{-\tau_v(s_0, s)} + \int_{s_0}^s B_v(T(s')) e^{-\tau_v(s', s)} ds'$$

We employed the Eddington approximation for inserting our seasonal-sampled remote explanatory covariates. The Eddington approximation is a special case of the two stream approximation commonly applied in spectral unmixing algorithms [www.esri.com]. The approximation assumed that the intensity in the georeferenced, canopied, riverine, blue-green algae LULC was a linear function of $\mu = \cos \theta$ (i.e., $I_v(\mu, z) = a(z) + \mu b(z)$)

where z was the normal direction to the medium. Note that expressing angular integrals in terms of μ simplified inserting the sampled covariates as $d\mu = -\sin\theta d\theta$ which appeared in the Jacobian of integrals as spherical coordinates.

Spatial linear predictors were generated from the canopy, shaded, blue-green algae LULC and the riffled water components using an ordinary kriged-based interpolator and traditional Landsat™ data. The dependent variable was the spectral emissivity estimates generated from the decomposition of the Landsat™ pixel which was transformed to fulfill the diagnostic normality test prior to performing kriging. The Ordinary kriged-based algorithm was then used to generate predictive, autoregressive, risk maps. Ordinary kriging was selected to interpolate the value $Z(x_0)$, the blue-green algae LULC and riffle

water components, $Z(x)$, at an unobserved bloom location x_0 from the field and remote-sampled covariate coefficient estimates and $z_i = Z(x_i)$, $i=1, \dots, n$ at nearby predicted riverine locations, x_1, \dots, x_n . Ordinary kriging is computed as a linear unbiased estimator, (\bar{x}_0) of $Z(x_0)$, based on a stochastic model of the dependence quantified by the variogram $\gamma(x, y)$ and by the expectation $\mu(x) = E[Z(x)]$ and the covariance function $c(x, y)$ of the random field [2]. The kriging estimator was given by a linear combination of the algorithm: using the canopy, shaded-related, endmember, spectral dataset of $z_i = Z(x_i)$ with weights $w_i(x_0)$, $i=1, \dots, n$ chosen such that the variance in the spectral autoregressive predictive model was calculated using which was further minimized similar to the interpolation in the model using multiple unmixing algorithms. and the ENVI classifiers.

Results and Discussion

A digital elevation and NDVI map of Ecuador illustrates the varied topographical features which should be taken into account to assess risk. In addition, freely available satellite imagery for this country focuses on the central region, which illustrates the utility of an interpolation signature method for areas in which satellite data may not be available. This simulation demonstrates that with the usage of concrete data – such as specific cholera case rates at the parish or district level, and enough satellite data to identify the presence and abundance of blue-green algae, the relationship between this type of vegetation and cholera incidence, and thus the potential for future outbreaks – can be assessed with precision.

Conclusion

In its 2012 epidemiological report on cholera in the year 2011, the WHO asserted the necessity to shift from cholera emergency response toward prevention, including a focus on behavior change [3]. However, due to the geographic and environmental conditions that favor bacterial growth – in addition to socioeconomic conditions – some communities are far more at risk of cholera than others. An interpolation technique has the potential to go a long way toward a seasonal detection system for this serious water-borne disease.

References

1. Malavade SS, Narvaez A, Mitra A, Ochoa T, Naik E, et al. (2011) Cholera in Ecuador: Current relevance of past lessons learnt. *J Global Infect Dis* 3: 189-194.
2. Tauxe RV, Mintz ED, Quick RE (1995) Epidemic cholera in the new world: translating field epidemiology into new prevention strategies. *Emerging Infectious Diseases* 1: 141.
3. WHO (2012) Cholera: Weekly epidemiological record, Geneva. Organisation mondiale de la Santé, Geneva, Switzerland.
4. Kumate J, Sepúlveda J, Gutiérrez G (1998) Cholera epidemiology in Latin America and perspectives for eradication. *Bull Inst Pasteur* 96: 217-226.
5. Fernandez MAL, Schomaker M, Mason PR, Fesselet JF, Baudot Y, et al. (2012) Elevation and cholera: an epidemiological spatial analysis of the cholera epidemic in Harare, Zimbabwe, 2008-2009. *BMC public health* 12: 442.
6. Ali M, Emch M, Donnay JP, Yunus M, Sack RB (2002) Identifying environmental risk factors for endemic cholera: a raster GIS approach. *Health place* 8: 201-210.
7. Chevallier E, Grand A, Azais JM (2004) Spatial and temporal distribution of cholera in Ecuador between 1991 and 1996. *Eur J Public Health* 14: 274-279.
8. Bates SJ, Trostle J, Cevallos WT, Hubbard A, Eisenberg JN (2007) Relating diarrheal disease to social networks and the geographic configuration of communities in rural Ecuador. *Am J Epidemiol* 166: 1088-1095.

-
9. de Magny GC, Murtugudde R, Sapiano MR, Nizam A, Brown CW, et al. (2008) Environmental signatures associated with cholera epidemics. *Proceedings of the National Academy of Sciences* 105: 17676-17681.
 10. Islam MS, Drasar BS, Sack RB (1994) Probable role of blue-green algae in maintaining endemicity and seasonality of cholera in Bangladesh: a hypothesis. *J Diarrhoeal Dis Res* 12: 245-256.
 11. Islam MS, Drasar BS, Bradley DJ (1990) Long-term persistence of toxigenic *Vibrio cholerae* 01 in the mucilaginous sheath of a blue-green alga, *Anabaena variabilis*. *J Trop Med Hyg* 93: 133-139.
 12. Epstein PR (1993) Algal blooms in the spread and persistence of cholera. *Biosystems* 31: 209-221.