academicJournals

Vol. 6(11), pp. 347-389, November 2014
DOI: 10.5897/JPHE2014.0629
ISSN 2006-9723
Article Number: 614874747911

http://www.academicjournals.org/JPHE

Journal of Public Health and Epidemiology

Full Length Research Paper

Denoising a model employing automated bandwidth selection procedures and pre-whitened Euclidean-based quadratic surrogates in PROC ARIMA for optimizing asymptotic expansions and simulations of onchocerciasis endemic transmission zones in Burkina Faso

Benjamin G. Jacob[1], Robert J. Novak[1], Laurent Toe[2], Moussas S. Sanfo[2], Rose Tingueria[4], Alain Pare[3], Mounkaila Noma[3], Daniel Griffith[4] and Thomas R. Unnasch[1]

[1]Department of Global Health, College of Public Health, University of South Florida, Tampa, FL, USA.
[2]Multi-Disease Surveillance Centre (MDSC), 1473 Avenue Naba Zombré - Ouagadougou, Burkina Faso.
[3]African Programs for Onchocerciasis Control (APOC), Epidemiology and Vector Elimination, 1473 Avenue Naba Zombré, Ouagadougou, Burkina Faso.
[4]United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Ouagadougou, Burkina Faso.
[5]School of Economic, Political and Policy Sciences. The University of Texas, Dallas, 800 West Campbell Road, Richardson, TX 75080-3021, USA.

Received 15 February, 2014; Accepted 22 August, 2014

In this research we constructed multiple predictive ArcGIS Euclidean distance–based autoregressive infectious disease transmission oriented models for predicting geographic locations of endemic onchocerciasis ("river blindness") transmission risk zones in Burkina Faso. We employed multiple spatiotemporal-sampled empirical ecological data sets of georeferenced covariates of riverine larval habitats of *Similium damnosum s.l.,* a black fly vector of onchocerciasis and their surrounding villages with their retrospective tabulated prevalence rates. The estimators were regressed employing the modified sum of squares technique. The model also revealed that 5 to 10 km was mesoendemic, 10 to 15 was hypoendemic and after 15 km there was no transmission. Semi-parametric spatial filtering matrices, orthogonal eigenvectors and interpolated endmember signatures can be used to render robust ARIMA risk model residual forecasts by reducing latent unobservable error coefficients in regressed spatiotemporal field-sampled immature *S. damnosum s.l.* density count data for optimizing risk mapping of seasonal onchocerciasis endemic transmission zones.

Key words: Autoregressive integrated moving average (ARIMA), QuickBird, *Similium damnosum s.l.*, onchocerciasis, Burkina Faso.

INTRODUCTION

In predictive autoregressive vector arthropod-related infectious disease transmission-oriented risk-based

statistics, ordinary least squares (OLS) would be amethod for estimating unknown seasonal parameter error estimators in a linear regression-based model (Griffith, 2005). The OLS is a method for estimating the unknown parameters in a linearized regression model (Hosmer and Lemeshew, 2000). This method would minimize the sum of squared vertical distances between the observed responses in the sampled dataset and the responses predicted by the linear approximation. The resulting estimator can then be expressed by a simple formula, especially in the case of a single regressor in a seasonal, vector, arthropod-related, endemic, transmission-oriented, explanatory model. The OLS estimator is consistent when the regressors are exogenous and there is no perfect multicollinearity, and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated (Cressie, 1993).

Optimally, thereafter, the class of linear, unbiased, seasonal, autoregressive, vector, arthropod-related, infectious disease, parameter estimators would be then homoscedastic and serially uncorrelated. A sequence or a vector of random variables is homoscedastic if all random variables in the sequence or vector have the same finite variance (Cressie, 1993). In a spatiotemporal, predictive, autoregressive, arthropod-related, risk-based, infectious disease distribution model, this would be known as homogeneity of variance (Mcdonald, 2008). Under these conditions, the method of OLS would provide minimum-variance, mean-unbiased, estimation when the endemic, transmission-oriented, regression-based, risk-based distribution model residual error coefficients have finite variances. Unfortunately, when employing an autoregressive prediction model for quantitating georeferenced, predictive, seasonal, vector, arthropod-related, spatially, error-prone, explanatory, predictor variables from past time-series, dependent, covariate, coefficient, indicator values for the forecasting equation must be initialized in order to fit the explanatory, observational predictors, employing OLS estimates for ascertaining robust inferences.

Fortunately, an eigenvector spatial filtering procedure can capture dependence based on the standard OLS estimator and is apart from the assumptions of independence and constant variance of the disturbances, a distribution freely owed to the Gauss Markov theorem. The Gauss Markov theorem states that in a linearized regression model in which the errors have expectation zero and are uncorrelated and have equal variances, the best linear, unbiased estimator of the sampled, covariate, coefficient indicator values would be provided by the OLS estimator *(*Aitken, 1935). This non-parametric spatial filtering approach can employ eigenvectors that are extracted from a transformed spatial link matrix for quantizing time-series, dependent, arthropod-related, autoregressive, infectious disease, transmission-oriented, risk-based, model residual autocorrelation error coefficients. The spatial, filtering estimator is fairly robust to endemic, transmission-oriented, model specification uncertainties compared with a spatial maximum likelihood estimator (McDonald, 2008). Commonly, assumptions of asymptotic properties of the maximum likelihood estimator and the quasi-maximum likelihood estimator have been employed for deducing parameter estimator significance in spatial autoregressive models (Cressie, 1993). For example, asymptotic expansions in a spatiotemporal, arthropod-related infectious disease, transmission-oriented risk-based stochastic interpolator would be represented as an increasing-domain asymptotic as it would be based on increasingly dense observational predictors in a fixed and bounded region.

In this research, we employed a space–time eigen decomposition spatial filter algorithm and multivariate autoregressive integrated moving average (ARIMA) for identifying onchocerciasis, endemic-oriented, transmission zones by quantitating a large, ecological, empirical dataset of georeferenced, riverine black fly larval habitat of *Similium damnosum s.l.* (that is, capture point) spatiotemporally-sampled in a study site in Burkina Faso.

Simuliidae or black flies in the *Simulium damnosum* Theobald complex are the only insect vectors of human onchoceriasis in West African countries (www.who.gov).ARIMA models are, in theory, the most general class of models for forecasting a time-series which can be stationarized by transformations such as differencing and logging. In fact, the easiest way to think of ARIMA models is as fine-tuned versions of random-walk and random-trend models: the fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors (Cressie 1993). In West Africa, members of *Simulium damnosum* complex are the only known vectors of human onchocerciasis (Crosskey 1960). Onchocerciasis a parasitic disease caused by infection by *Onchocerca volvulus*, a nematode (roundworm) which is the world's second-leading infectious cause of blindness. The parasite is transmitted to humans through the bite of a blackfly of the genus *Simulium* commonly found in fast flowing rural river ecosystems. *Similium* larval stages are commonly found in running water where Precambrian rocks break the water surface and the turbulence of the water results in a higher level of oxygenation (Crosskey 1960).

Our assumption was that spatial dependence in the *S.*

*Corresponding author. E-mail: bjacob1@.health.usf.ed.

damnosum s.l. riverine larval habitat in the data was from unobservable latent predictor variables that were correlated. Spatial dependence is the existence of statistical dependence in a collection of random variables or a collection time series of random variables, each of which is associated with a different geographical location (Goodchild, 1980). This dependence is naturally formulated within the framework of hierarchical, spatiotemporal, arthropod-related, infectious, disease, parameter estimator models and over the past decade, a variety of spatial models have been proposed for quantitating the latent level(s) of the hierarchy in these models. This is because dependence is of prime impor-tance in these applications where it is quite reasonable to postulate the existence of corresponding set of random variables at specific endemic transmission zones (for example, hyperendemic) that have not been included in a sample. Unfortunately, the specific issues posed by the sparseness of competent error estimators in predictive algorithms for arthropod-related infectious disease related data for quantitating local spatial dependence have not been thoroughly addressed in literature.

As such, initially, in this research, a georeferenced immature *S. damnosum s.l*. riverine habitat capture point was overlaid onto QuickBird visible and near-infra-red (NIR) (www.digitalglobe.com) data based on stratified geographical locations at 5, 10 and 15 km Euclidean-based distances from a capture point. Numerous studies have been undertaken using satellite-derived environmental data to predict the distribution, abundance and prevalence of diseases and their vectors, including malaria (Hay et al., 2000; Rogers et al., 2002), leishmaniasis (Elnaiem et al., 2003), filariasis (Lindsay and Thomas, 2000), trypanosomiasis (Rogers, 2000) and schistosomiasis (Brooker et al., 2001; Malone et al., 2001; Brooker et al., 2002a, 2002b; Moodley et al., 2003; Kabatereine et al., 2004). Univariate and Poisson regression models were then constructed for each delineated transmission zone. Thereafter, an autoregressive approach was employed to spatially extrapolate the existence of any residualized stochastic processes in the mean of the regression models.

Euclidean, distant-based, explanatory measurements were generated in ArcGIS spatial analyst from the georeferenced *S. damnosum s.l.,* riverine, larval habitat capture point which was then employed to delineate the endemic, transmission-oriented zones thresholds at the study site, employing stratified prevalence rates as independent variables in the regression-based estimation matrices. The asymptotic distribution of an empirical dataset of georeferenced, standardized, linear model parameter estimators were then derived to detect if the serially correlated latent processes were present in the ArcGIS Euclidean distant-based measurements. This technique qualitatively assessed the time series dependence in the spatiotemporal-sampled dataset of explanatory, predictor, covariate, coefficient estimates which were then log-transformed into Gaussian independent variables. These covariate coefficients were then exported into an autoregressive, uncertainty, probabilistic estimation model framework in SAS/GIS. Currently, within SAS/GIS the ARIMA procedure provides parameter estimation for constructing autoregressive integrated moving average (that is, Box-Jenkins) models, seasonal ARIMA models, transfer function models and intervention models (www.sas.edu).

Prior to mapping the onchocerciasis-related variables, exploratory spatial data analysis (ESDA) tools in ArcGIS were used to assess the statistical properties of the field-sampled data. Having explored the data, we created a variety of output map types (that is, prediction, error of prediction, probability and quantile) using a variant of a stochastically-based, kriged-based, explanatory algorithm (that is an, ordinary interpolator) and associated tools (for example, data transformation, declustering and detrending). Our research also considered the construction of specific parameter estimators of regression coefficients in a linear regression model, employing stochastically oriented a priori information. A priori information can be framed as stochastic restrictions (Cressie, 1993). In this research, the dominance conditions of the estimators were derived under the criterion of mean squared error matrix. Simple probabilistic and disjunctive formulas for quantitating the effect of the estimated predictive, residual, standard, autocovariance error variables were then generated. Thereafter, we adjusted the bias in eigenspace using spatiotemporal-sampled Euclidean, distance-based parameter estimators for deriving precise endemic transmission zones (that is, 5, 10 and 15 km) as depicited by the ArcGIS delineated maps created from the georeferenced, riverine *S. damnosum s.l.* riverine larval habitat capture point. We employed spatiotemporal-sampled data obtained from the African Programme for Onchocerciasis Control (APOC, 1974–2002) for remotely constructing our robust, endemic,*S. damnsoum s.l,.* riverine topographic, riverine –based landscape forecasting risk models. Large scale control of onchocerciasis commenced over three decades ago, initially through the Onchocerciasis Control Programme in West Africa (OCP, 1974–2002), and more recently by the African Programme for Onchocerciasis Control (APOC, 1995–2010). The goals of OCP were to eliminate onchocerciasis as a public health problem and to mitigate its negative impact on the social and economic development of affected regions (Toe, 1993). The strategic objective of APOC is to permanently protect the remaining 120 million people at risk of this debilitating and disfiguring disease in 19 countries in Africa through the establishment of community-directed treatment with ivermectin (CDTI) that is capable of being sustained by the communities after APOC financing has ended.

Additionally, in this research, robust, predictive endmember signatures were generated from a spectrally decomposed georeferenced, *S. damnosum s.l.,* riverine

Figure 1. Map of Burkina Faso.

larval habitat capture point using QuickBird mixel data, employing multiple un-mixing models and object-based classifiers (for example, Li –Strahler geometric –optical model, ENVI's Spectral Angular Mapper). Because of the design specifications of sensors, rarely, if ever does the spatial resolution match the size of an item on the ground; when one pixel includes the signatures of two or more endmembers, it is considered a mixed pixel or a 'mixel' (Jensen, 2005). In this research, the endmembers derived were stochastically interpolated for identifying unknown, unsampled riverine, larval habitats along the study site corridor. The analyses also included the spatial-spectral, endmember extraction algorithm (SSEE), which was employed using the SPA. Our assumption was that by interpolating unmixed, sub-meter, resolution, sub-mixel, image, riverine, larval, habitat, capture point endmember, emissivity spectra extracted from various unmixing algorithms in a least squares estimation algorithm, the residual would reveal unsampled *S damnosum s.l.,*riverine larval habitats and their within canopied features (for example, Precambrian rock and ripple water). Spectral unmixing algorithms have proliferated in a variety of ecological disciplines by exploiting remotely sensed data (Jensen, 2005).

Therefore, this research objectives were to: (1) Remotely display all spatiotemporal, seasonal-sampled, empirical empirical, ecological-based, *S. damnosum s.l.*- related spatial feature attributes with their surrounding riverine–based communities using QuickBird visible and near infra-red (NIR) data; (2) construct multiple predictive, autoregressive models employing time series-dependent, explanatory, covariate coefficients; (3) spectrally extract and decompose a QuickBird mixel to derive and classify endmember emissivity spectra for interpolating a target signature; (4) construct residual uncertainty covariance matrices based on regression-derived observational predictors and; (4) adjust any bias in the Euclidean distance–based parameter estimators at distances of 5, 10 and 15 km from a capture point to generate a robust, autoregressive, endemic, transmission-oriented, predictive risk map, delineating onchocerciasis-related endemic transmission zones at a georeferenced, epidemiological riverine study site in Burkina Faso.

METHODOLOGY

Study site

Burkina Faso is a landlocked country in West Africa. It is surrounded by six countries: Mali to the north, Niger to the east, Benin to the southeast, Togo and Ghana to the south, and Côte d'Ivoire to the southwest (Figure 1) . Its size is 274,200 km^2 (105,900 sq. m)

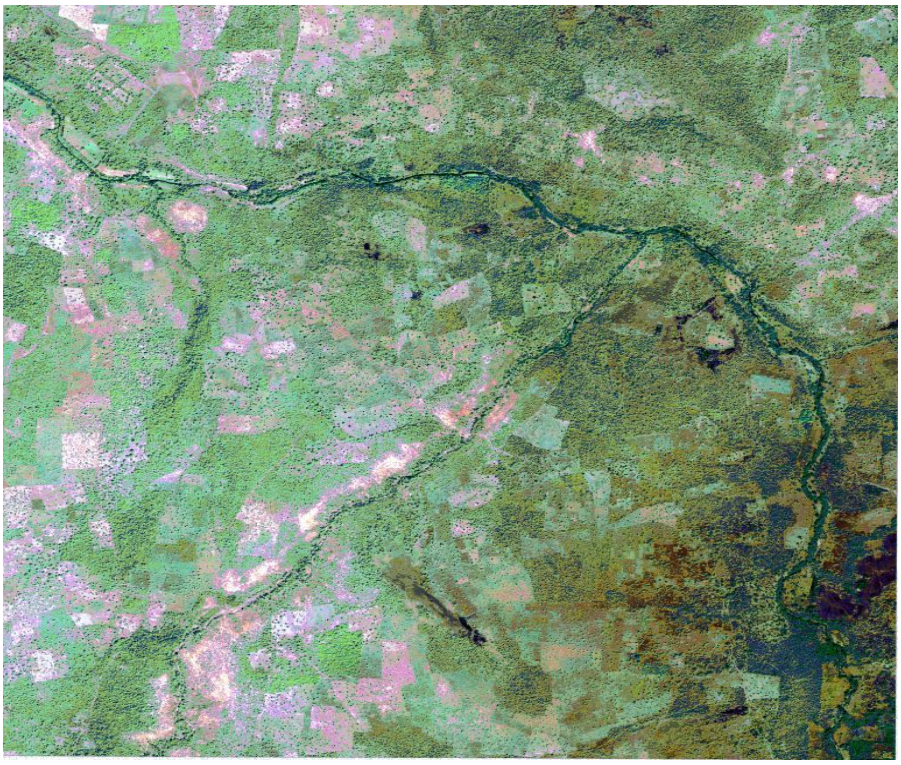


Figure 2. QuickBird visible and near infra-red data for the Chutes-Dienkoa study site.

with an estimated population of more than 15,757,000. Total land area is 274,200 km² of which water covers approximately 400 km². Burkina Faso has three distinct seasons: warm and dry (November to March), hot and dry (March to May), and hot and wet (June to October). Annual rainfall varies from about 250 mm to 1,000 mm in the riverine study site region. The terrain is mostly flat, with undulating plains and hills. Most of the study site region lies on a savanna plateau, with fields, brush and scattered tree. Burkina Faso lies mostly between latitudes 9° and 15°N (a small area is north of 15°), and longitudes 6° W and 3° E. It is made up of two major types of countryside. The larger part of the country is covered by a peneplain, which forms a gently undulating landscape with, in some areas, a few isolated hills, the last vestiges of a Precambrian Massif. The Southwest of the country, on the other hand, forms a sandstone massif, where the highest peak, Ténakourou, is found at an elevation of 749 m (2,457 ft). The massif is bordered by sheer cliffs up to 150 m (492 ft) high. The average altitude of Burkina Faso is 400 m (1,312 ft) and the difference between the highest and lowest terrain is no greater than 600 m (1,969 ft). Burkina Faso is therefore a relatively flat country. The country owes its former name of Upper Volta to three rivers which crosses it: the Black Volta the White Volta and the Red Volta. The Black Volta is one of the country's only two rivers which flow year-round, the other being the Komoé, which flows to the southwest. The basin of the Niger River also drains 27% of the country's surface. The Niger's tributaries – the Béli, the Gorouol, the Goudébo and the Dargol are seasonal streams and flow for only four to six months a year.

Remote sensing data

Raster image data from the DigitalGlobe QuickBird satellite service were acquired for the study site for the periods of 15th July, 2010, within the riverine study site area (Figure 2). In this research the QuickBird image data were delivered as pan-sharpened composite products in infra-red (IR) colors. QuickBird multispectral products provided four discrete non-overlapping spectral bands in the 0.45 to 0.72 µm range. The QuickBird sensors were able to identify dug wells that were 1 to 2 cm in depth. Results revealed that well-digging was practiced on 387 (1.4%) rainfed land cover classified areas, 15,638 (54.7%) with the majority located in dryer arid regions. The field-plot revealed an accuracy of 92% with an error of omission and commission of less than 10%. Only the clearest, cloud-free imagery was available of the contiguous sub-areas of the study site. The Order Polygon contained 5 vertices consisting of longitude/latitude (decimal degrees) geographic coordinates using a WGS-84 ellipsoid. The satellite data contained 64 km^2 of the land cover in the riverine epidemiological study site. The QuickBird imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS *Imagine* v.8.7™ (ERDAS, Inc., Atlanta, Georgia). A base map of the riverine study site was then generated in ArcGIS using the QuickBird data and differentially corrected global positioning systems (DGPS) ground coordinates of the spatiotemporal-sampled *S. damnsoum s.l.* habitat epidemiological capture point and the surrounding georeferenced villages (Figure 3).

The DGPS were acquired from a CSI max receiver which has a positional accuracy of +/- 0.178. (http://www.omnistar.com/). Using a local DGPS broadcaster can compensate for ionospheric and ephemeris effects which can improve horizontal accuracy significantly and can bring altitude error down in a predictive vector insect habitat model (Jensen, 2005). Each georeferenced *S.damnosum s.l.* habitat was entered into the VCMS™ relational database software product (Clarke Mosquito Control Products, Roselle, IL). The VCMS™ database supports a mobile field data acquisition component module called Mobile VCMS™ that synchronizes field-sampled arthropod-related data from industry standard Microsoft Windows Mobile™ devices and can support

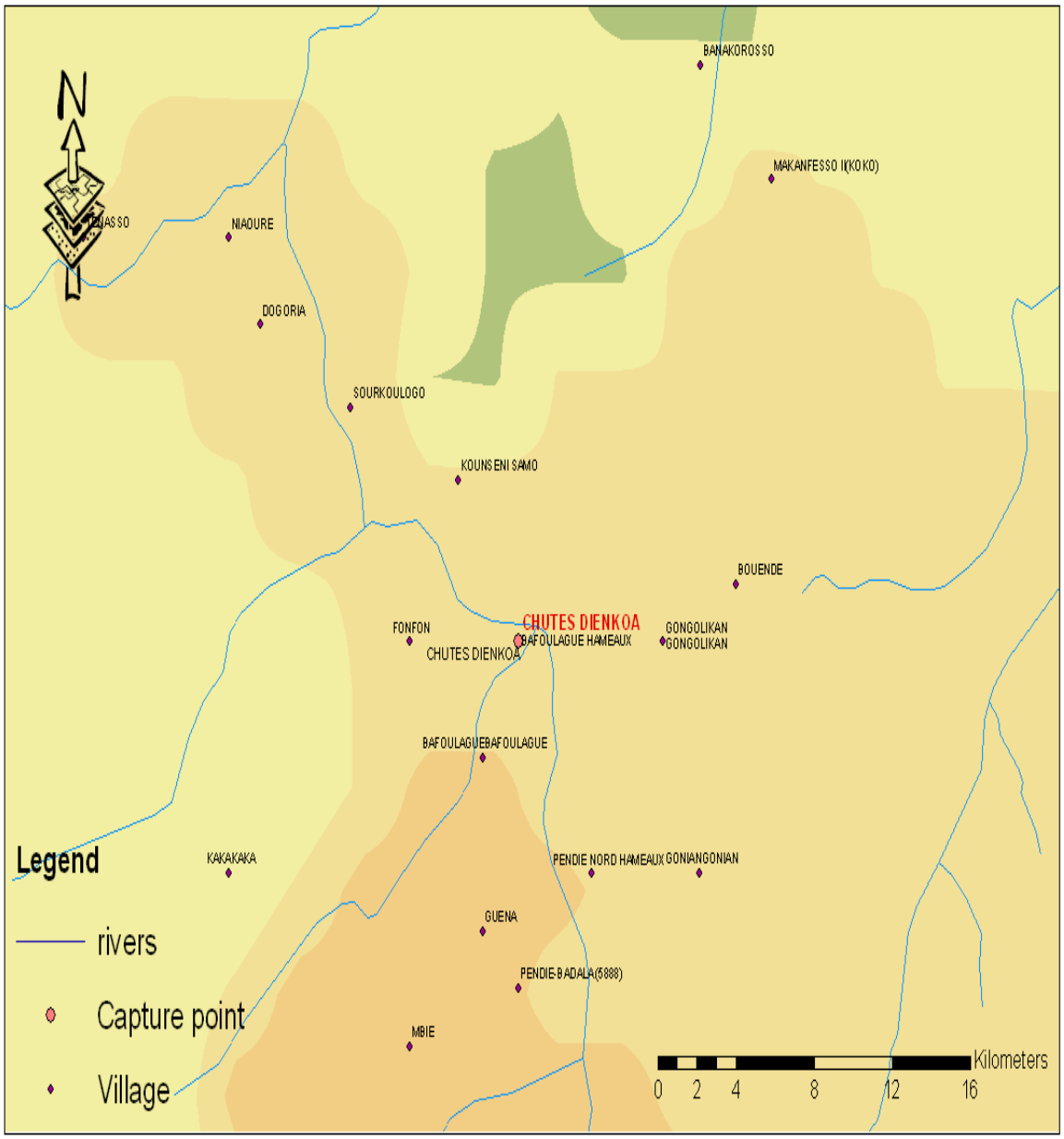


Figure 3. Base map of the sampled study site capture point and surrounding epidemiological sampled riverine villages.

add-on DGPS data collection (Jacob et al., 2008b, c). A digitized grid-based algorithm was then constructed in ArcGIS by applying a mathematical algorithm in order to fit the continuous and bounded sampled larval habitat surfaces from a field-sampled attribute. A grid is a raster data storage format native to ESRI (www.esri.com) (Figure 3).

Environmental parameters

Multiple observational georeferenced explanatory observational predictors were then examined extensively using longitude, latitude, and altitude data (Table 1). The criteria involved the centrographic measures of spatial mean and distance between the epidemiological prevalence stratified villages and the georeferenced capture point. The data was also comprised of individual spatiotemporal-sampled observations of *S. damnosums.l.* habitat capture point together with a battery of categorical attribute measures which were expanded to include multiple, endemic, transmission-oriented, risk-based, explanatory, predictor, covariate, coefficient estimates.

The habitat distances were then measured as Euclidean distances in the ArcGIS projection units of the raster which computed 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 georeferenced epidemiological villages and from the riverine capture point to the villages (that is, geometric

Table 1. Environmental predictor variables sampled at the epidemiological capture point.

Variable	Description	Units
GCP	Ground control points	Decimal-degrees
FlOW	flowing water	Presence or absence
HGHT	Height of water	-
TURB	Turbidity of water	Formazin Turbidity Unit
AQVEG	Aquatic vegetation	Percentage
HGVEG	Hanging vegetation	Percentage
DDVEG	Dead vegetation	Percentage
RCKS	Rocks	Percentage
MMB	Man-made barriers	Type (damns, bridges)
DISHAB	Distance between habitats	Meters

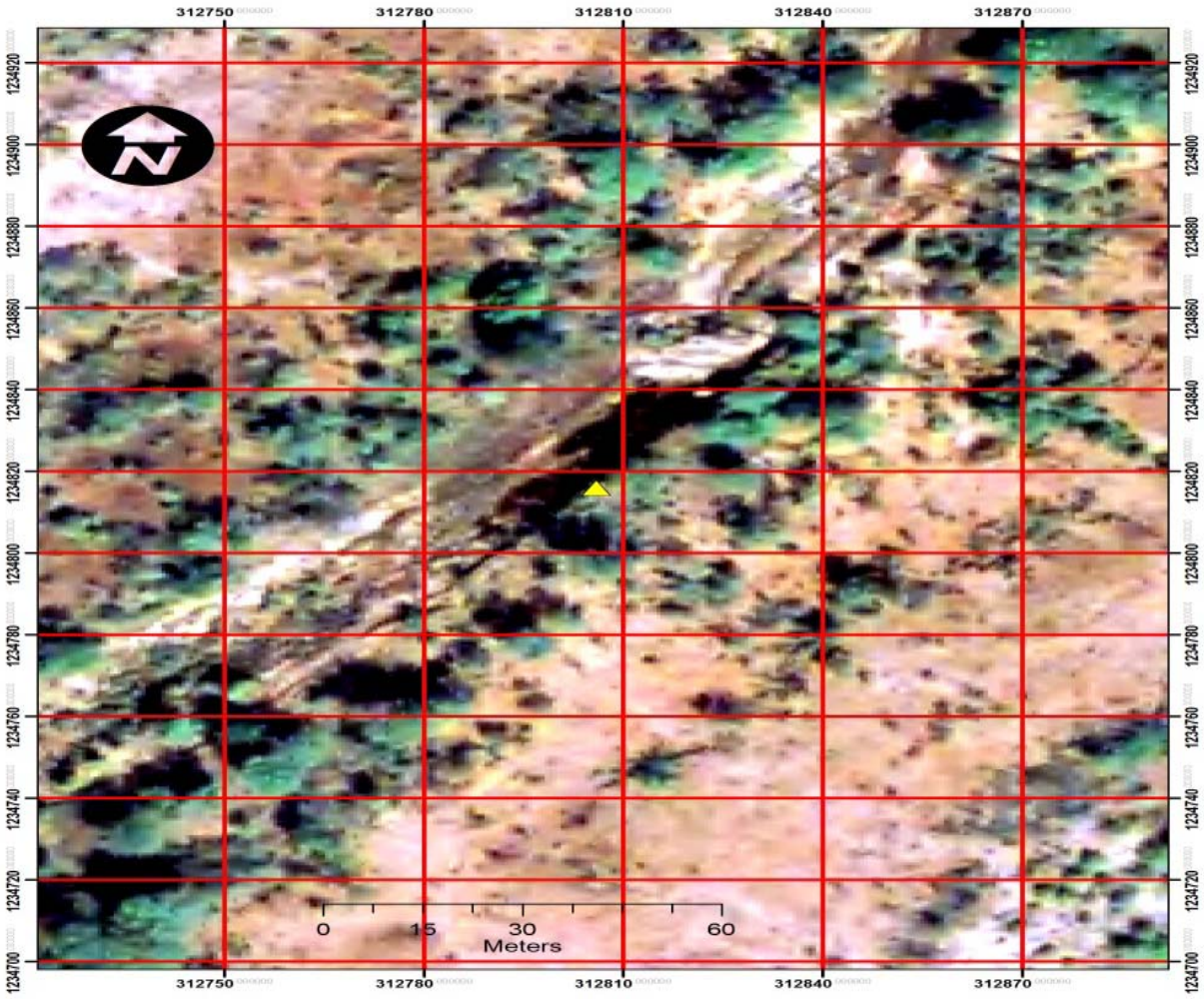


Figure 4. An ArcGIS digitized grid matrix overlaid onto the georeferenced *S. damnosum s.l.* riverine habitat capture point.

distances in the multidimensional space). In this research, the Euclidean distances were computed as: distance (x,y) = $\{ \sum_i (x_i - y_i)^2 \}^{½}$. Every cell in the Euclidean allocation output raster was the then assigned the value of the source to which it was closest. The nearest source was then determined by the Euclidean distance function in ArcGIS®. This function assigned space between the georeferenced *S. damnosums.l.* riverine habitat capture point and the villages with their stratified prevalence rates. The Euclidean direction output raster contained the Azimuth direction from each grid cell centroid to the nearest source. Then the Euclidean allocation function identified the nearest human habitation center closest to each grid cell. The distance between sampled and human habitation areas were then categorized into numerous classes (for example, 1: 0 to 5 km, 2: 5 to 10 km and 10 to 15 km) Figure 4.

Regression model

Poisson models were constructed for determining significance levels of the Euclidean distance-based explanatory covariates and the other endemic transmission-oriented LULC attributes in SAS GEN MOD. The Poisson process was provided by the limit of a binomial distribution of the spatiotemporal-sampled, covariate, coefficient estimates within each distance boundary set at 5 km intervals from the capture point up to a maximum distance of 15 km employing:

$$P_p(n \mid N) = \frac{N!}{n!\,(N-n)!}\, p^n (1-p)^{N-n}, \quad (1).$$

Prevalence rates were used as the response variable in the models. We viewed the distribution as a function of the expected number of e predictor variables sampled employing the sample size N for quantifying the fixed p in equation (1), which then was transformed into the linear equation:

$$P_{\gamma \mid N}(n \mid N) = \frac{N!}{n!\,(N-n)!}\left(\frac{\gamma}{N}\right)^n \left(1-\frac{\gamma}{N}\right)^{N-n},$$

Based on the sample size N, the distribution approached $P_\gamma(n)$ in this research was expressed as:

$$\lim_{N\to\infty} P_p(n \mid N) = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!}\,\frac{\gamma^n}{N^n}\left(1-\frac{\gamma}{N}\right)^N \left(1-\frac{\gamma}{N}\right)^{-n} = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n}\,\frac{\gamma^n}{n!}\left(1-\frac{\gamma}{N}\right)^N \left(1-\frac{\gamma}{N}\right)^{-n}$$

$$= 1\cdot\frac{\gamma^n}{n!}\cdot e^{-\gamma}\cdot 1 = \frac{\gamma^n e^{-\gamma}}{n!},$$

The GENMOD procedure then was used to fit multiple generalized linear models (GLMs) equations to the sampled riverine epidemiological data by maximum likelihood estimation of the parameter vector β. In this research the GENMOD procedure estimated the explanatory covariate coefficients of each Euclidean distance-based model at 5, 10 and 15 km numerically through an iterative fitting process. The dispersion parameter was then estimated by the residual deviance and by employing the product of Pearson's chi-square divided by the degree of freedom (d.f.) in the model. Covariances, standard errors, and *p*-values were then computed for the estimated explanatory, observational, predictor at each distant-dependent, geographical location based on the asymptotic normality derived from the maximum likelihood estimation. Note that the sample size N completely dropped out of the probability function, which in this research had the same functional form for all the spatiotemporal-sampled, distant-dependent parameter estimator values (that is, γ). As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1 (Haight, 1967). The ratio of probabilities was then:

$$\sum_{n=0}^{\infty} P_\gamma(n) = e^{-\gamma}\sum_{n=0}^{\infty}\frac{\gamma^n}{n!} = e^{-\gamma}e^{\gamma} = 1.$$ provided by: $$\frac{P_\gamma(n=i+1)}{P(n=i)} = \frac{\frac{\gamma^{i+1}e^{-\gamma}}{(i+1)!}}{\frac{e^{-\gamma}\gamma^i}{i!}} = \frac{\gamma}{i+1}.$$

The Poisson distribution revealed that the covariate coefficients reached a maximum when: $$\frac{dP_\gamma(n)}{dn} = \frac{e^{-\gamma}n(\gamma - H_n + \ln\gamma)}{n!} = 0,$$

where γ was the Euler-Mascheroni constant and H_n was a harmonic number, leading to the transcendental equation: $\gamma - H_n + \ln\gamma = 0$.

The model revealed that the Euler-Mascheroni constant arose in the integrals as:

$$\gamma = -\int_0^\infty e^{-x}\ln x\,dx = -\int_0^1 \ln\ln\left(\frac{1}{x}\right)dx = \int_0^\infty\left(\frac{1}{1-e^{-x}}-\frac{1}{x}\right)e^{-x}\,dx = \int_0^\infty\frac{1}{x}\left(\frac{1}{1+x}-e^{-x}\right)dx \quad (2).$$

Commonly, integrals that render γ in combination with temporal constants include:

$$\int_0^\infty e^{-x^2}\ln x\,dx = -\frac{1}{4}\sqrt{\pi}\,(\gamma+2\ln 2)$$ and $$\int_0^\infty e^{-x}(\ln x)^2\,dx = \gamma^2+\frac{1}{6}\pi^2,$$ (Haight 1967).

Thereafter, the double integrals in the distant-based regression models included:

$$\gamma = \int_0^1\int_0^1 \frac{x-1}{(1-xy)\ln(xy)}\,dx\,dy$$

An interesting analog of equation (2) in the models was then provided by:

$$\ln\left(\frac{4}{\pi}\right)_{=}\sum_{n=1}^{\infty}(-1)^{n-1}\left[\frac{1}{n}-\ln\left(\frac{n+1}{n}\right)\right]_{=}\int_0^1\int_0^1\frac{x-1}{(1+xy)\ln(xy)}\,dx\,dy_{=0.241564\ldots\gamma.}$$

This solution was also provided by incorporating Mertens theorem:

$$e^{\gamma}=\lim_{n\to\infty}\frac{1}{\ln p_n}\prod_{i=1}^{n}\frac{1}{1-\frac{1}{p_i}},$$

[i.e.,]

where the product was aggregated over the sampled explanatory, predictor,covariate, coefficient values found in the ecological datasets. Mertens' 3rd theorem:

$$\lim_{n\to\infty}\ln n\prod_{p\leqslant n}\left(1-\frac{1}{p}\right)=e^{-\gamma},$$

is related to the density of prime numbers, where γ is the Euler–Mascheroni constant. By taking the logarithm of both sides in the regression models, an explicit formula for ?l was then dervived using:

$$\gamma=\lim_{x\to\infty}\left[\sum_{p\leqslant x}\ln\left(\frac{1}{1-\frac{1}{p}}\right)-\ln\ln x\right].$$

This product was also given by series due to Euler, which followed from equation (2) by first replacing:

$\ln n_{\text{b}}\ln(n+1)$, in the equation $\gamma=\sum_{k=1}^{\infty}\left[\frac{1}{k}-\ln\left(1+\frac{1}{k}\right)\right]$ and then generating $\lim_{n\to\infty}[\ln(n+1)-\ln n]=\lim_{n\to\infty}\ln\left(1+\frac{1}{n}\right)=0,$.

We then substituted the telescoping sum: $\sum_{k=1}^{n}\ln\left(1+\frac{1}{k}\right)$ for $\ln(n+1)$ which rendered $\ln\left(1+\frac{1}{k}\right)=\ln(k+1)-\ln k,$.

Thereafter, we obtained: $\lim_{n\to\infty}\left[\sum_{k=1}^{n}\frac{1}{k}-\sum_{k=1}^{n}\ln\left(1+\frac{1}{k}\right)\right]_{?=}\lim_{n\to\infty}\sum_{k=1}^{n}\left[\frac{1}{k}-\ln\left(1+\frac{1}{k}\right)\right]$

Additionally, other series in the distant-based regression models included the equation (∩) where:

$$?=\sum_{n=2}^{\infty}(-1)^n\frac{\zeta(n)}{n}=\ln\left(\frac{4}{\pi}\right)+\sum_{n=1}^{\infty}\frac{(-1)^{n-1}\zeta(n+1)}{2^n(n+1)}$$

and where $\zeta(z)$ was $\gamma=\sum_{n=1}^{\infty}(-1)^n\frac{\lfloor \lg n\rfloor}{n}$ and the Riemann zeta function. The Riemann zeta functionζ(s), is a function of a complex, explanatory, predictor variables that analytically continues the sum of the infinite series: $\sum_{n=1}^{\infty}\frac{1}{n^s},$

which converges when the real part of *s* is greater than 1 where lg is the logarithm to base 2 and $\lfloor x\rfloor$ is the floor function (see Haight 1967).

In this research $\binom{n}{k}$ was employed as a binomial coefficient, and then it was rearranged to achieve the conditionally convergent series in our Euclidean, distant-based, endemic, transmission-oriented, regression-based, epidemiological, predictive, risk model as the plus and minus terms were first grouped in pairs of the sampled covariate coefficients using the resulting series of the actual seasonal-sampled values. The double series was thereby equivalent to Catalan's integral:

$$\gamma = \int_0^1 \frac{1}{1+x} \sum_{n=1}^{\infty} x^{2^n-1}\, dx.$$

[i.e.,]

Catalan's integrals are a special case of general formulas due to:

$$J_0\left(\sqrt{z^2-y^2}\right) = \frac{1}{\pi}\int_0^{\pi} e^{y\cos\theta}\cos(z\sin\theta)\, d\theta,$$

where $J_0(z)$ is a Bessel function of the first kind. The Bessel function is a function $Z_n(x)$ defined in a robust regression model by employing the recurrence relations:

$$Z_{n+1} + Z_{n-1} = \frac{2n}{x} Z_n \quad \text{and} \quad Z_{n+1} - Z_{n-1} = -2\frac{dZ_n}{dx}.$$

solutions and quantify heteroskedastic parameters in a spatiotemporal regression models using the differential equation: $x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(x^2 - n^2\right)y = 0.$ (Haight 1967), which more recently has been employed to define (Ross, 2007).

In this research the Bessel function $J_n(z)$ was defined by the contour integral:

$$J_n(z) = \frac{1}{2\pi i}\oint e^{(z/2)(t-1/t)}\, t^{-n-1}\, dt,$$

where the contour enclosed the origin and was traversed in a counterclockwise direction. This function generated:

$$J_0\left(2i\sqrt{z}\right) = \frac{1}{\pi}\int_0^{\pi} e^{(1+z)\cos\theta}\cos\left[(1-z)\sin\theta\right] d\theta.$$

where $z = 1 - z'$ and $y = 1 + z'$. Thereafter, to quantify the equivalence in the sampled empirical dataset of the regression-based parameter estimators, we expanded $1/(1+x)$ in a geometric series and multiplied the district-level sampled data by x^{2^n-1}, and integrated the term wise as in Sondow and Zudilin (2006). Other series for γ then included:

$$\gamma = \frac{3}{2} - \ln 2 - \sum_{m=2}^{\infty}(-1)^m \frac{m-1}{m}\left[\zeta(m) - 1\right] \quad \text{and} \quad \gamma = \frac{2^n}{e^{2^n}}\sum_{m=0}^{\infty}\frac{2^{mn}}{(m+1)!}\sum_{t=0}^{m}\frac{1}{t+1} - n\ln 2 + O\left(\frac{1}{2^n e^{2^n}}\right),$$

A rapidly converging limit for γ was then provided by: $\gamma = \lim_{n\to\infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n}\left(\frac{1}{k} - \frac{\zeta(1-k)}{n^k}\right)\right] = \lim_{n\to\infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n}\frac{1}{k}\left(1 + \frac{B_k}{n^k}\right)\right],$ where B_k was a Bernoulli number.

Another limit formula was then provided by the equation:

$$\gamma = -\lim_{n\to\infty}\left[\frac{\Gamma\left(\frac{1}{n}\right)\Gamma(n+1)\, n^{1+1/n}}{\Gamma\left(2+n+\frac{1}{n}\right)} - \frac{n^2}{n+1}\right]$$

Limits to the model was then rendered by:

$$\gamma = \lim_{x\to\infty}\zeta(\zeta(z)) - 2^x + \left(\frac{4}{3}\right)^x + 1$$

where $\zeta(z)$ was the Riemann zeta function. Another model output was rendered from the endemic transmission-oriented regression-based equation which was generated by $d(n) = \sigma_0(n)$ for the linearly, quantitated, explanatory, covariate coefficient numerical values in the empirical dataset 1 to n in the sampled dataset which in this research was found to be asymptotic to:

$$\frac{\sum_{k=1}^{n} d(k)}{n} \sim \ln n + 2\gamma - 1$$

An elegant identity for γ in our regression models was then provided by:

$$\gamma = \frac{S_0(z) - K_0(z)}{I_0(z)} - \ln\left(\frac{1}{2}z\right),$$

Where $I_0(z)$ was a modified Bessel function of the first kind, $K_0(z)$ was a modified Bessel function of the second kind, and:

$$S_0(z) = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}z\right)^{2k} H_k}{(k!)^2},$$

where H_n was a harmonic number. This provided an efficient iterative algorithm for γ by computing:

$$B_k = \frac{B_{k-1}\, n^2}{k^2}, A_k = \frac{1}{k}\left(\frac{A_{k-1}\, n^2}{k} + B_k\right), U_k V_k = U_{k-1} + A_k \text{ and } V_k = V_{k-1} + B_k \text{with } A_0 = -\ln n B_0 = 1 U_0 = A_0 \text{ and } V_0 = 1.$$

Reformulating this identity rendered the limit in our endemic, transmission-oriented, regression-based,risk model as:

$$\lim_{n\to\infty}\left[\sum_{k=0}^{\infty} \frac{\left(\frac{n^k}{k!}\right)^2 H_k}{\sum_{k=0}^{\infty}\left(\frac{n^k}{k!}\right)^2} - \ln n\right] = \gamma.$$

In this research, infinite products involving γ also arose from the Barnes G-function using the sampled, explanatory, covariate coefficients. In mathematics, the Barnes G-function $G(z)$ is a function that is an extension of super factorials to the complex numbers which is related to the Gamma function. In this research, this function provided:

$$\prod_{n=1}^{\infty} e^{-1+\frac{1}{2n}}\left(1+\frac{1}{n}\right)^n = \frac{e^{1+\frac{\gamma}{2}}}{\sqrt{2\pi}} \text{ and } \prod_{n=1}^{\infty} e^{-2+\frac{2}{3n}}\left(1+\frac{2}{n}\right)^n = \frac{e^{3+2\gamma}}{2\pi}.$$

The Barnes G-function was thereafter linearly defined in our time-series, dependent,endemic, transmission-oriented distribution, risk model which was then defined by using the product of:

$$G(z+1) = (2\pi)^{z/2} \exp\left(-\left(z(z+1)+\gamma z^2\right)/2\right) \times \prod_{n=1}^{\infty}\left[\left(1+\frac{z}{n}\right)^n \exp\left(-z + z^2/(2n)\right)\right],$$

Where γ was the Euler–Mascheroni constant, $\exp(x) = e^x$, and ∏ was the capital pi notation. The Euler-Mascheroni constant was thereafter rendered by the expressions $\gamma = -\Gamma'(1) = -\psi_0(1)$, where $\psi_0(x)$ was the digamma function:

$$\gamma = \lim_{s\to 1^+}\left[\zeta(s) - \frac{1}{s-1}\right] \text{ and the symmetric limit form of: } \gamma = \lim_{s\to 1^+}\sum_{n=1}^{\infty}\left(\frac{1}{n^s} - \frac{1}{s^n}\right) \text{ and } \gamma = \lim_{x\to 0}\left[x - \Gamma\left(\frac{1}{x}\right)\right]$$

The digamma function was then defined as the logarithmic derivative of the gamma function: $\psi(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}.$

where it was the first of the polygamma functions in the endemic, transmission-oriented, regression-based, risk-related model. The digamma function, often denoted as $\psi_0(x)$, $\psi^0(x)$ is related to the harmonic numbers in a seasonal, infectious disease, arthropod-related, risk model in that $\psi(n) = H_{n-1} - \gamma$ where H_n is the n^{th} harmonic number, and γ is the Euler-Mascheroni constant (Jacob et al., 2012a). For half-integer values, the digamma function may be expressed as: $\psi\left(n+\frac{1}{2}\right) = -\gamma - 2\ln 2 + \sum_{k=1}^{n}\frac{2}{2k-1}$ (Hosmer and Lemeshow, 2000).

The digamma function in our model was denoted as $\psi_0(x)$ which was related to the harmonic numbers in $\psi(n) = H_{n-1} - \gamma$ when H_n was the n^{th} harmonic number, and γ was the Euler-Mascheroni constant. It had the integral representation:

$$\psi(x) = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1-e^{-t}} \right) dt$$

In this research this expression was written as: $\psi(s \mid 1) = -\gamma \mid \int_0^1 \frac{1-x^s}{1-x} dx$ which followed from Euler's integral formula for the harmonic numbers derived from the linear endemic transmission regression-based risk-related model. In this research the Digamma of the linearized *S. damnosum s.l.* endemic-transmission-oriented risk-related model was computed in the complex plane, using:

$$\psi(z+1) = -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n+z)} \qquad z \neq -1, -2, -3, \ldots$$

and

$$\psi(z) = -\gamma + \sum_{n=0}^{\infty} \frac{z-1}{(n+1)(n+z)} = -\gamma + \sum_{n=0}^{\infty} \left(\frac{1}{n+1} - \frac{1}{n+z} \right) \qquad z \neq 0, -1, -2, -3, \ldots$$

These equations were utilized to evaluate infinite sums of rational functions, where $p(n)$ and $q(n)$ were polynomials of n. Performing partial fraction on u_n in the complex field, in the spatiotemporal, predictive, autoregressive, vector, arthropod-related, endemic, transmission-oriented, landscape, risk-based, distribution models employing all roots of $q(n)$ as simple roots then rendered:

$$u_n = \frac{p(n)}{q(n)} = \sum_{k=1}^{m} \frac{a_k}{n+b_k}.$$

We had to use $\lim_{n \to \infty} n u_n = 0,$ for the series to converge. Hence, the expression:

$$\sum_{k=1}^{m} a_k = 0, \text{ and } \sum_{n=0}^{\infty} u_n = \sum_{n=0}^{\infty} \sum_{k=1}^{m} \frac{a_k}{n+b_k} = \sum_{n=0}^{\infty} \sum_{k=1}^{m} a_k \left(\frac{1}{n+b_k} - \frac{1}{n+1} \right) = \sum_{k=1}^{m} \left(a_k \sum_{n=0}^{\infty} \left(\frac{1}{n+b_k} - \frac{1}{n+1} \right) \right) = -\sum_{k=1}^{m} a_k \left(\psi(b_k) + \gamma \right) = -\sum_{k=1}^{m} a_k \psi(b_k).$$

was rendered using the series expansion of higher rank polygamma function and a generalized formula:

[i.e., $\sum_{n=0}^{\infty} u_n = \sum_{n=0}^{\infty} \sum_{k=1}^{m} \frac{a_k}{(n+b_k)^{r_k}} = \sum_{k=1}^{m} \frac{(-1)^{r_k}}{(r_k-1)!} a_k \psi^{(r_k-1)}(b_k)$].

We noticed that in our endemic transmission-oriented model ψ was the only solution of the functional equation: $F(x+1) - F(x) + \frac{1}{x}$,that is monotone on $\mathbb{R}^+$ satisfied $F(1) = -\gamma$. The digamma then had a Gaussian sum of the form:

$$\frac{-1}{\pi k} \sum_{n=1}^{k} \sin\left(\frac{2\pi n m}{k} \right) \psi\left(\frac{n}{k} \right) = \zeta\left(0, \frac{m}{k} \right) = -B_1\left(\frac{m}{k} \right) = \frac{1}{2} - \frac{m}{k}.$$

Thereafter, the endemic, transmission-oriented, risk-based, distribution, model residuals revealed that $\zeta(s,q)$ was the Hurwitz zeta function and $B_n(x)$ which was a Bernoulli polynomial. The Bernoulli polynomials are an Appell sequence with $g(t) = \frac{e^t - 1}{t}$ (Roman 1984), giving the generating function $\frac{t e^{tx}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x) \frac{t^n}{n!}$ (Abramowitz and Stegun 1972), first obtained by Euler (1738). The first few Bernoulli polynomials in this research were:

$B_0(x) = 1$, $B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}$, $B_3(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x$, $B_4(x) = x^4 - 2x^3 + x^2 - \frac{1}{30}$, $B_5(x) = x^5 - \frac{5}{2}x^4 + \frac{5}{3}x^3 - \frac{1}{6}x$, $B_6(x) = x^6 - 3x^5 + \frac{5}{2}x^4 - \frac{1}{2}x^2 + \frac{1}{42}$.

We also defined an older type of "Bernoulli polynomial" by writing:

$$t\frac{e^{xt}-1}{e^t-1}=\sum_{n=1}^{\infty}\frac{\phi_n(z)\,t^n}{n!}$$

This would then render the polynomials $\phi_n(x)=B_n(x)-B_{n+1}$ where B_n was a Bernoulli number, the first few of which are:

$$\phi_1(x)=x,\phi_2(x)=x^2-x,\phi_3(x)=x^3-\frac{3}{2}x^2+\frac{1}{2}x,\phi_4(x)=x^4-2x^3+x^2,\phi_5(x)=x^5-\frac{5}{2}x^4+\frac{5}{3}x^3-\frac{1}{6}x.$$

The Bernoulli polynomials also satisfied $B_n(1)=(-1)^n B_n(0)$ and $B_n(1-x)=(-1)^n B_n(x)$ for $n\neq 1, B_n(1)=B_{n+1}$ so $B_n(1)=B_n=0$ for odd $n\geq 1$. The *S. damnosum s.l.* polynmoials also satisfied the relation $B_n(x+1)-B_n(x)=n\,x^{n-1}$ In this research, for deriving robust estimation values of x, $B_n(x)$ was expressed for the explanatory, predictor, covariate coefficients in the empirical dataset integers *n*in terms of Bernoulli and Euler numbers which led to:

$$B_n(1)=(-1)^n B_n,\ B_n\left(\tfrac{1}{2}\right)=\left(2^{1-n}-1\right)B_n,\ B_n\left(\tfrac{1}{4}\right)=-2^{-n}\left(1-2^{1-n}\right)B_n-4^{-n}\,n\,E_{n-1},\ B_{2n}\left(\tfrac{1}{3}\right)=-\tfrac{1}{2}\left(1-3^{1-2n}\right)B_{2n}$$

$$,\ B_{2n}\left(\tfrac{1}{6}\right)=\tfrac{1}{2}\left(1-2^{1-2n}\right)\left(1-3^{1-2n}\right)B_{2n}.$$

Bernoulli (1713) defined the polynomials in terms of sums of the powers of consecutive integers:

$$\sum_{k=0}^{m-1}k^{n-1}=\frac{1}{n}\left[B_n(m)-B_n(0)\right].$$

Fortunately, our Bernoulli polynomials satisfied the recurrence relation:

$$\frac{dB_n}{dx}=n\,B_{n-1}(x)$$

in the*S. damnosum s.l.*model which obeyed the identity $B_n(x)=(B+x)^n$, where B^k was interpreted as the Bernoulli number [i.e.,$B_k=B_k(0)$] (Jacob et al., 2005b).When formulated as an equation to be solved, recurrence relations are known as recurrence equations, or sometimes difference equations (Everitt 2002). The difference between the n^{th} convergent and γ in the onchocericais, endemic, transmission-oriented, linearized,regression-based, risk model was then provided by:

$$\sum_{k=1}^{n}\frac{1}{k}-\ln n-\gamma=\int_n^{\infty}\frac{x-\lfloor x\rfloor}{x^2}\,dx,$$

where $\lfloor x\rfloor$ was the floor function which satisfied the inequality expression: $\frac{1}{2(n+1)}<\sum_{k=1}^{n}\frac{1}{k}-\ln n-\gamma<\frac{1}{2n}$.

The symbol γ was then $\gamma'=e^{\gamma}\approx 1.781072$. This led to the radical representation of the seasonal-sampled, explanatory, predictor, covariate coefficients as:

$$e^{\gamma}=\left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/3}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/4}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/5}\cdots,$$ which was then related to the double series:

$$\gamma=\sum_{n=1}^{\infty}\frac{1}{n}\sum_{k=0}^{n-1}(-1)^{k+1}\binom{n-1}{k}\ln(k+1)$$ and $\binom{n}{k}$, a binomial coefficient.

Another proof of product in the regression models was then provided by the equation:

$$\frac{\pi}{2}=\left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/4}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/8}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/16}\cdots.$$

The solution was then made clearer by changing $n\to n+1$. In this research both these regression-based formulas were also analogous to the product for $e$which was then rendered by the computation:

$$e = \left(\frac{2}{1}\right)^{1/1}\left(\frac{2^2}{1\cdot 3}\right)^{1/2}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/3}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/4}\ldots$$

Unfortunately, extra-Poisson variation was detected in the estimated predictive, residual variance estimates in our model. As such, we constructed a robust negative binomial regression model in SAS with non-homogenous gamma distributed n means by incorporating $\alpha = \frac{1}{\theta}(\alpha > 0)$ in equation (2.1) .The distribution in the linear regression was then rewritten:

$$f(y_i|x_i) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}, \quad y_i = 0, 1, 2, \ldots$$

The negative binomial distribution was thus derived as a gamma mixture of the Poissonian randomized variables. The conditional mean in the model was then $E(y_i|x_i) = \mu_i = e^{x_i'\beta}$ and the variance was:

$$V(y_i|x_i) = \mu_i[1 + \frac{1}{\theta}\mu_i] = \mu_i[1 + \alpha\mu_i] > E(y_i|x_i)$$

To further estimate the district-level models, we specified DIST=NEGBIN (p=1) in the MODEL statement in PROC REG. The negative binomial model NEGBIN1 was set $p = 1$, which revealed the variance function [i.e., $V(y_i|x_i) = \mu_i + \alpha\mu_i$] was linear in the mean of the models. The log-likelihood function for each NEGBIN1 regression model was then provided by the following equation:

$$\mathscr{L} = \sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln\left(j + \alpha^{-1}\exp(x_i'\beta)\right)\right.$$

$$\left. - \ln(y_i!) - \left(y_i + \alpha^{-1}\exp(x_i'\beta)\right)\ln(1+\alpha) + y_i\ln(\alpha)\right\}.$$

where

The gradient for the risk model was then:

$$\frac{\partial\mathscr{L}}{\partial\beta} = \sum_{i=1}^{N}\left\{\left(\sum_{j=0}^{y_i-1}\frac{\mu_i}{(j\alpha+\mu_i)}\right)x_i - \alpha^{-1}\ln(1+\alpha)\mu_i x_i\right\}$$ and

$$\frac{\partial\mathscr{L}}{\partial\alpha} = \sum_{i=1}^{N}\left\{-\left(\sum_{j=0}^{y_i-1}\frac{\alpha^{-1}\mu_i}{(j\alpha+\mu_i)}\right) - \alpha^{-2}\mu_i\ln(1+\alpha) - \frac{(y_i+\alpha^{-1}\mu_i)}{1+\alpha} + \frac{y_i}{\alpha}\right\}.$$

In this research, the negative binomial regression district-level model with variance function $V(y_i|x_i) = \mu_i + \alpha\mu_i^2$, was referred to as the NEGBIN2 model. To estimate this model, we specified DIST=NEGBIN (p=2) in the MODEL statements. A test of the Poisson distribution was then performed by examining the hypothesis that $\alpha = \frac{1}{\theta} = 0$. A Wald test of this hypothesis was also provided which were the reported t statistics for the estimates in the negative binomial regression model. The log-likelihood function of the model (that is, NEGBIN2) was then generated by

$$\mathscr{L}n = \sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln(j+\alpha^{-1}) - \ln(y_i!) - (y_i+\alpha^{-1})\ln(1+\alpha\exp(x_i'\beta)) + y_i\ln(\alpha) + y_i x_i'\beta\right\}$$ = where y was an integer when the gradient was:

$$\frac{\partial\mathscr{L}}{\partial\beta} = \sum_{i=1}^{N}\frac{y_i-\mu_i}{1+\alpha\mu_i}x_i.$$

The variance in the model was then assessed by: $$\frac{\partial\mathscr{L}}{\partial\alpha} = \sum_{i=1}^{N}\left\{-\alpha^{-2}\sum_{j=0}^{y_i-1}\frac{1}{(j+\alpha^{-1})} + \alpha^{-2}\ln(1+\alpha\mu_i) + \frac{y_i-\mu_i}{\alpha(1+\alpha\mu_i)}\right\}.$$

Object-oriented classification

Once an ecological dataset of the remotely-dependent, explanatory, predictor, covariate coefficients was constructed in ArcGIS, the data was exported to ENVI® which used various spectral–based algorithms to analyze the QuickBird visible and NIR data of the georeferenced *S. damnosums.l.*riverine larval habitat capture point.

The two main algorithms employed in our endmember decomposition was the spectral angle mapper (SAM) and spectral information divergence (SID) classification. SAM is a deterministic method that looks for an exact pixelmatch and weights the differences as same while SID is a probabilistic method that allows for variations in pixel measurements, where probability is measured from zero to a user-defined threshold.

In our research the basic workflow involved importing the data collected in the field from the riverine study site into a spectral library (http://www.exelisvis.com). Thereafter, SAM employed an *n*-dimensional angle to match the QuickBird pixels to the reference spectra. The algorithm determined the spectral similarity between the spectra by calculating the angle between the spectra and treating them as vectors in a space with dimensionality equal to the number of satellite bands. SAM compared the angle between the endmember spectrum vector and each QuickBird pixel spectrum vector in *n*-D space. Smaller angles represented in the dataset revealed closer matches to the reference spectrum. Pixels further away than the specified maximum angle threshold in radians were not classified. SID is a spectral classification method that uses a divergence measure to match pixels to reference spectra (http://www.exelisvis.com). The smaller the divergence, the more likely the pixels are similar. Pixels with a measurement greater than the specified maximum divergence threshold are not classified. 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 QuickBird image (Figure 5).

Additionally, in this research we used the Sequential Maximum Angle Convex Cone (SMACC) spectral tool in ENVI to determine the spectral endmembers and their abundances throughout the image. SMACC is designed to use a convex cone model (that is, residual minimization) to identify image endmember spectra (http://www.exelisvis.com). Extreme points were used to determine a convex cone, which defined the first *S. damnsoum s.l.* riverine larval habitat endmembers in the dataset. A constrained oblique projection QuickBird was then applied to the existing cone to derive the next larval habitat endmember. The cone was increased to include new endmembers. The process was repeated until a projection derived at endmember that already existed within the convex cone (to a specified tolerance) or until the specified number of endmembers were found.

The image endmembers of the georeferenced, *S.damnosum s.l.,*riverine, larval habitat and its associated spatial, date, feature, attributes were then extracted them from ENVI®'s spectral library. Several spectra corresponding to the different backgrounds in the sampled, canopy-oriented, capture, point structures (that is, Precambrian rock and rippled water pixel components) had to be included, since multiple scatterings between floating leaves in the habitat, for example, and a bright soil background increased the QuickBird NIR reflectance. Leaf cells have evolved to scatter (that is, reflect and transmit) solar radiation in the NIR spectral region (Schowengerdt, 1997). After calibration, the spectrally defined covariate coefficient estimates from the image were converted to match the library. Analogously, the QuickBird reference endmembers spectra in the library was transformed into the endmembers spectra of the image. Image classification was then performed, employing the FLAASH™ object-oriented approach which rendered a gmd file that converted the image's digital number (DN) to at-sensor radiance and computed at-sensor reflectance while normalizing the solar elevation angle. The equation was as follows:

$$\rho_{BandN} = \frac{\pi(L_{BandN} * Gain_{BandN} + Bias_{BandN}) * D^2}{E_{BandN} * (COS((90-\theta) * \pi / 180))}$$ where,

ρ_{BandN} = Reflectance for Band N
L_{bandN} = DN for Band N
D = Normalized Earth-Sun Distance
E_{bandN} = Solar Irradiance for Band N

Spectral decomposition

A predictive 3-Dimensional (D) radiative transfer equation employing the sampled *S. damnosum s.l.* riverine larval habitat spatial,data, feature attributes was constructed. In order to characterize larval, habitat, hotspot phenomenon effectively and obtain stable solutions of canopy, multiple scattering, the radiation field was decomposed into three parts; unscattered radiance

[i.e., $\mathrm{I}^0(\tau,\Omega)$],

single scattering radiance

[i.e., $\mathrm{I}^1(\tau,\Omega)$],

and the multiple scattering radiance

[i.e., $\mathrm{I}^{\mathrm{M}}(\tau,\Omega)\,\mathrm{I}(\tau,\Omega) = \mathrm{I}^0(\tau,\Omega) + \mathrm{I}^1(\tau,\Omega) + \mathrm{I}^{\mathrm{M}}(\tau,\Omega)$].

A simple scheme was then represented by $\mathrm{I}^0(\tau,\Omega)$ which in this research was denoted by 1, which was neither scattered by the atmosphere nor canopy, but was reflected directly by the canopy surface. In this research, $\mathrm{I}^1(\tau,\Omega)$ was radiance either scattered once by the atmosphere, denoted by 2, or once by the canopy, denoted by the value 3. Also $\mathrm{I}^{\mathrm{M}}(\tau,\Omega)$ was the most complicated spectral component, which included all of other riverine larval, habitat, canopy components in the radiation field of the coupled medium. Unscattered sunlight radiances $\mathrm{I}^0(\tau,\Omega)$ were then characterized by the following radiative transfer equation and corresponding boundary conditions. When T < Ta the *S. damnosum s.l.,*riverine, larval habitat, radiative transfer model rendered:

$$\begin{cases} -\mu\dfrac{\partial \mathrm{I}^0(\tau,\Omega)}{\partial\tau} + \mathrm{I}^0(\tau,\Omega) = 0 \\ \mathrm{I}^0(0,\Omega) = \delta(\Omega-\Omega_0)i_0 \qquad \mu < 0 \\ \mathrm{I}^0(\tau_\alpha^{bot},\Omega) = \mathrm{I}^0(\tau_c^{top},\Omega) \qquad \mu > 0 \end{cases}$$

where τ_a^{bot} and τ_c^{top} were the optical depths at the bottom of the atmosphere and the top of the riverine, larval, habitat canopy, respectively. Here different notations were employed to indicate the physical meaning of the canopy boundary condition. The model

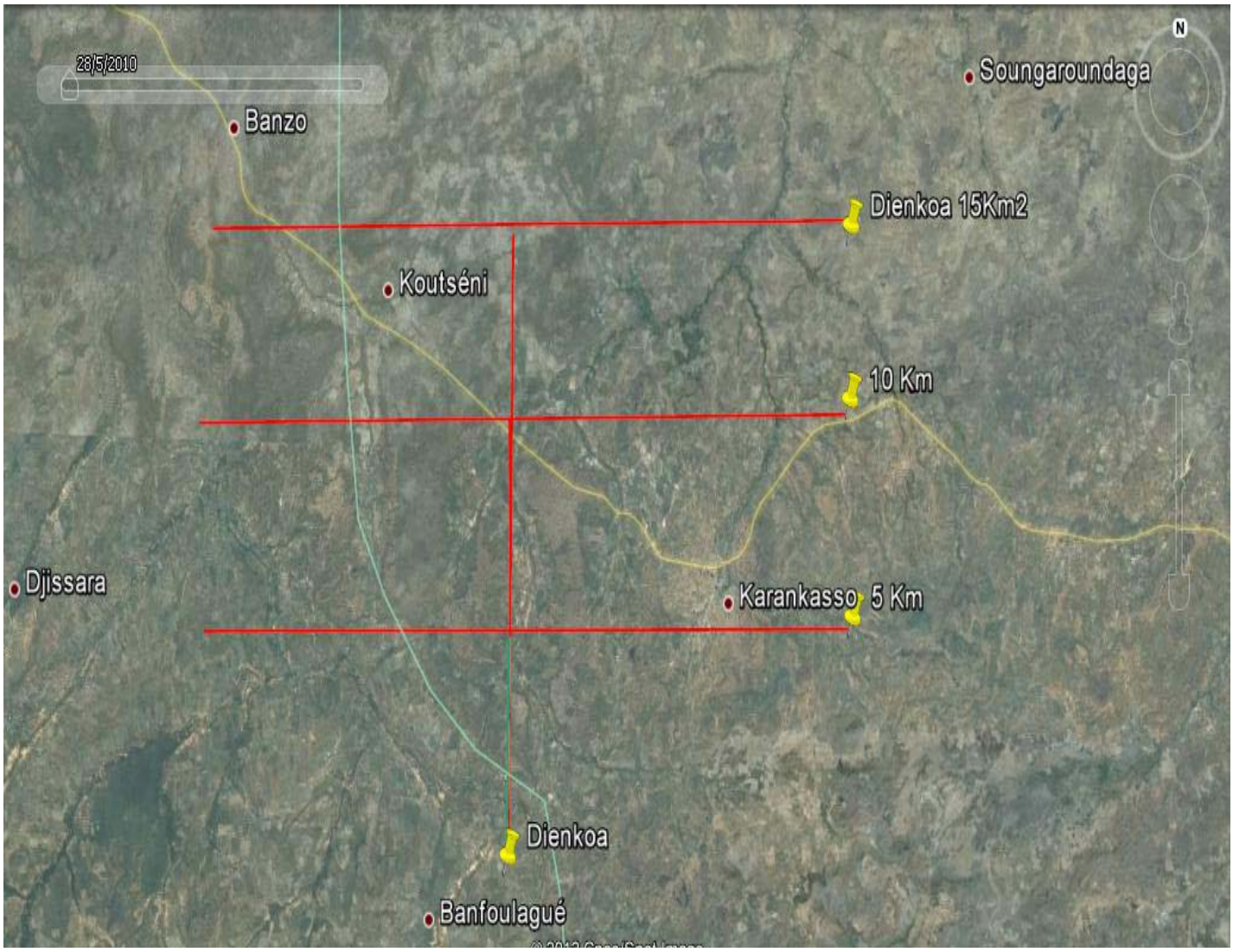


Figure 5. A Euclidean-distance matrix overlaid onto the georeferenced *S. damnosum s.l.* riverine habitat capture point.

provided the upper boundary condition, which meant only parallel sunlight illuminated the atmosphere at the top in the direction Ω_0. When $\tau > \tau_\alpha$, the residuals were:

$$\begin{cases} -\mu \dfrac{\partial \mathrm{I}^0(\tau,\Omega)}{\partial \tau} + h(\tau,\Omega)G(\Omega)\mathrm{I}^0(\tau,\Omega) = 0 \\ \mathrm{I}^0\left(\tau_c^{top},\Omega\right) = \mathrm{I}^0\left(\tau_\alpha^{bot},\Omega\right) \qquad \mu < 0 \\ \mathrm{I}^0(\tau_t,\Omega) = fs(\Omega_0,\Omega)|\mu_0|\mathrm{I}^0(\tau_t,\Omega_0) \qquad \mu > 0 \end{cases}$$

Jointly solving the equations with these boundary conditions, it was easy to obtain

$$
\mathrm{I}^{0}(\tau,\Omega)=
\begin{cases}
\mathrm{I}^{0}_{d1}(\tau,\Omega)=i_0\exp\left(-\tau/|\mu|\right)\delta(\tau,\Omega_0) & \mu<0,\tau\le\tau_\alpha \\
\mathrm{I}^{0}_{d2}(\tau,\Omega)=\mathrm{I}^{0}_{d1}(\tau,\Omega) & \\
\quad\cdot\exp\left[-C(\Omega)(\tau-\tau_\alpha)/|\mu|\right] & \mu<0\tau_\alpha<\tau\le\tau_t \\
\mathrm{I}^{0}_{u2}(\tau,\Omega)=i_0 & \\
\quad\cdot\exp\left[-\dfrac{\tau_\alpha+(\tau_t-\tau_\alpha)G(\Omega_0)}{|\mu_0|}\right] & \\
\cdot\ fs(\Omega_0,\Omega)|\mu_0|\exp\left[-\varepsilon(\tau,\Omega)\right] & \mu>0,\tau_\alpha<\tau\le\tau_t \\
\mathrm{I}^{0}_{u1}(\tau,\Omega)=\mathrm{I}^{0}_{u2}(\tau_\alpha,\Omega) & \\
\cdot\exp\left[-(\tau_\alpha-\tau)/\mu\right] & \mu>0,\tau\le\tau_\alpha
\end{cases}
$$

where $\mathrm{I}^{0}_{u2}(\tau,\Omega)$ represented the upwelling sunlight radiance within the georeferenced canopied *S. damnosum s.l.* riverine larval habitat capture point, and the function $\varepsilon(\tau,\Omega)$, due to modifying the extinction coefficient of the canopy. We then incorporated the extracted canopy radiance values including the Precambrian rock and ripple water spectral components using:

$$
\begin{aligned}
\varepsilon(\tau,\Omega)&=\frac{1}{\pi}\int_{\tau}^{\tau_t}h(t,\Omega)G(\Omega)dt \\
&=G(\Omega)\frac{\tau_t-\tau}{\mu}-\left[\sqrt{\frac{G(\Omega_0)G(\Omega)}{\mu|\mu_0|}}\frac{k\mathrm{H}}{\Delta(\Omega_0,\Omega)}\right]t_0
\end{aligned}
$$

Where t0 was defined as $t_0=\exp\left[-\dfrac{\Delta(\Omega_0,\Omega)\tau}{k\mathrm{H}}\right]-\exp\left[-\dfrac{\Delta(\Omega_0,\Omega)\tau_t}{k\mathrm{H}}\right]$

Since for single scattering radiances, unscattered sunlight becomes the scattering source (Kimes, 1991). In this research, boundary conditions were determined based on the fact that no incident single scattering radiances where from above the top of atmosphere or below the bottom of the canopy. When T < Ta the model rendered:

$$
\begin{cases}
-\mu\dfrac{\partial\mathrm{I}^{l}(\tau,\Omega)}{\partial\tau}+\mathrm{I}^{l}(\tau,\Omega)= & \\
\dfrac{\omega i_0}{4\pi}p(\Omega_0\rightarrow\Omega)\exp\left(-\dfrac{\tau}{\mu_0}\right) & \\
r^{1}(0,\Omega)=0 & \mu<0 \\
\mathrm{I}^{1}\left(\tau_\alpha^{bot},\Omega\right) & \mu>0
\end{cases}
$$

Thereafter, when T > T the georeferenced *S.damnosum s.l.* riverine larval hábitat capture point spectral decomposition rendered:

$$\begin{cases} -\mu\dfrac{\partial\tau^1(\tau,\Omega)}{\partial\tau}+h(\tau,\Omega)G(\Omega)\mathrm{I}^1(\tau,\Omega) \\ \dfrac{i_0'}{\pi}\Gamma(\Omega_0,\to\Omega)\exp\left[-(\tau-\tau_\alpha)\dfrac{G(\Omega_0)}{|\mu_0|}\right] \\ \mathrm{I}^1(\tau_c^{top},\Omega)=\mathrm{I}^1(\tau_\alpha^{bot},\Omega) \qquad \mu<0 \\ \mathrm{I}^1(\tau_t,\Omega)=0 \qquad \mu>0 \end{cases}$$

Where i_0' was the incident solar net flux arriving at the top of the riverine larval habitat canopy when $i_0'=i_0\exp(-\tau_\alpha/|\mu_0|)$

In the downward direction $\mu<0$, the solution was easily derived. When T < Ta, the spectral decomposition model was solved using:

$$\mathrm{I}^{l}(\tau,\Omega)=$$

$$\begin{cases} \dfrac{\omega Fop(\Omega_0\to\Omega)|\mu_0|}{4(|\mu_0|\leftarrow|\mu|)}\left[\exp\left(-\dfrac{\tau}{|\mu_0|}\right)-\exp\left(-\dfrac{r}{|\mu|}\right)\right] & \Omega\neq\Omega_0 \\ \dfrac{\omega F_0\tau}{4|\mu_0|}P(\Omega_0\to\Omega)\exp\left(-\dfrac{\tau}{\mu_0}\right) & \Omega=\Omega_0 \end{cases}$$

When $\tau_\alpha<\tau<\tau_t$ the model was solved using the equation:

$$\mathrm{I}^1(\tau,\Omega)=$$

$$\begin{cases} \dfrac{i_0'|\mu_0|\Gamma(\Omega_0\to\Omega)}{\pi\left[G(\Omega)|\mu_0|-G(\Omega_0)|\mu|\right]}t_1+\Delta\mathrm{I}^1(\tau,\Omega) & \Omega\neq\Omega_0 \\ \dfrac{(\tau-\tau_\alpha)i_0'\Gamma(\Omega_0\to\Omega)}{\pi|\mu_0|}\exp\left[-G(\Omega_0)\dfrac{\tau-\tau_\alpha}{|\mu_0|}\right] \\ \qquad +\Delta\mathrm{I}^1(\tau,\Omega) \end{cases}$$

Where $t_1$iwas defined by the equations:

$$t_1=\exp\left[-G(\Omega_0)\frac{\tau-\tau_\alpha}{|\mu_0|}\right]-\exp\left[-G(\Omega)\frac{\tau-\tau_\alpha}{|\mu|}\right]\quad\text{and}\quad \Delta\mathrm{I}^1(\tau,\Omega)=\mathrm{I}^1(\tau_\alpha,\Omega)\exp\left[-G(\Omega)(\tau-\tau_\alpha)/|\mu|\right.$$

which was the single scattering riverine larval habitat capture point canopy radiance emerging from the atmosphere without further scattering in the canopy. In the upward direction (p > 0), the solutions were a little more complicated because of the hotspot effect:

$$\mathrm{I}^1(\tau,\Omega)=\begin{cases}\dfrac{1}{\mu}\displaystyle\int_r^{\tau_t}F(\tau',\Omega)\\ \cdot\exp\left[-\dfrac{1}{\mu}\displaystyle\int_\tau^{\tau'}h(\varepsilon,\Omega)G(\Omega)d\varepsilon\right]d\tau' & \tau_\alpha\le\tau\le\tau_t\\ \dfrac{\omega F_0\,p(\Omega_0\to\Omega)|\mu_0|}{4(\mu+|\mu_0|)}t_2\\ +\mathrm{I}^1(\tau_\alpha,\Omega)\exp\left(\dfrac{\tau-\tau_\alpha}{\mu}\right) & \tau<\tau_\alpha\end{cases}$$

Where t_2 was defined using: $t_2=\exp\left[-\dfrac{\tau}{|\mu_0|}\right]-\exp\left[\dfrac{\tau}{\mu}-\left(\dfrac{1}{\mu_0}+\dfrac{1}{\mu}\right)\tau_\alpha\right]$ and the second integration as Ta < T < Tt which in this research was explicitly obtained by means of an alternative intergrand range. This range was solved using:

$$F(\tau',\Omega)=\frac{i_0'}{\pi}\Gamma(\Omega_0\to\Omega)\exp\left[-G(\Omega_0)(\tau-\tau_\alpha)/|\mu_0|\right]$$

.

The radiance $\mathrm{I}^1(\tau,\Omega)$ at Ta < T < Tt needed to be numerically evaluated without further assumptions. An explicit approximation to $\mathrm{I}^1(\tau,\Omega)$ was then derived and used for inversion.

In the spectral *S.damnosum s.l.*rivrerine larval habitat capture point endmember model the Gausse-Legendre quadrature was also employed to calculate the integration. An *n*-point Gaussian quadrature rule is a quadrature rule constructed to yield an exact result for polynomials of degree 2*n* − 1 or less by a suitable choice of the points x_i and weights w_i for i = 1,...,*n*. The domain of integration for such a rule is conventionally taken as [−1, 1], so the rule is stated as

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i).$$

We we used the Li-Strahler geometric-optical model based on the assumption that the BRDF would retrieve *S. damnosum s.l.* habitat capture point shaded riverine canopy structural variables. The BRDF was defined by:

$$f_r(\omega_i,\omega_o)=\frac{dL_r(\omega_o)}{dE_i(\omega_i)}=\frac{dL_r(\omega_o)}{L_i(\omega_i)\cos\theta_i\,d\omega_i}$$

Where *L* was the radiance, *E* was the irradiance, and θ_i was the angle made between ω_i and the riverine habitat and its associated Precambrian rock and ripple water surface reflectance emissivities. Because the BRDF is a four-dimensional function that defines how light is reflected at an opaque surface (Jensen, 2005), the function in our model took an incoming light direction, ω_i, and outgoing direction, ω_o, which were both defined with respect to the georeferenced *S. damnosum* riverine larval habitat and its neighboring Precambrian rock and ripple water surface *n*, and returned the ratio of reflected radiance exiting along ω_o to the irradiance incident from direction ω_i. Note, each direction ω was itself parameterized by azimuth angle φ and zenith angle θ, therefore, in this research, the BRDF was 4-dimensional. The BRDF had units sr^{-1}, with steradians (sr) being a unit of solid angle.

The inverted Li-Strahler geometric-optical model was then used to retrieve specific spectral habitat explanatory predictor covariate coefficient estimates. The reflectance associated with a georeferenced habitat was treated as an area-weighted sum of four fixed radiance components: sunlit canopy, sunlit background, shaded canopy, and shaded background. In most arthropod-related, infectious disease, larval, habitat–related, geometric-optical, simulation models these four components could be simplified to three: sunlit canopy–C, sunlit background–G and shadow–T (Jacob et al., 2011c). In this research, the endmember spectral components were derived using G, C, T components' classes which were initially estimated by the QuickBird image using ENVI®. For inverting the model, parts of the three spectral components were then represented by (kg) which was calculated using:

$$K_g=e^{-\pi\cdot M\cdot[\sec(\theta_i)+\sec(\theta_v)-O(\theta_i,\theta_v,\varphi)]} \quad (3)$$

$$O(\theta_i,\theta_v,\varphi)=1/\pi\left(\sec\theta_i+\sec\theta_v\right)(t-\sin t\cos t) \quad (3)$$

$$\cos t=\frac{h\left|\tan\theta_i-\tan\theta_v\cos\varphi\right|}{r\left(\sec\theta_i+\sec\theta_v\right)} \quad (4)$$

$$M=\frac{-\ln(K_g)}{\left(\sec\theta_i+\sec\theta_v\right)(\pi-t+\cos t\sin t)} \quad (5)$$

$$CC=1-e^{-\pi\cdot M} \quad (6)$$

Where, *I q u q* were the zenith angles of illumination and viewing, *O* was the average of the overlap function between illumination and viewing shadows of the capture point and their associated Precambrian rock and ripple water spectral components as

projected onto the background. In this research, *j* was the difference in azimuth angle between illumination and viewing.

In our analyses, the BRDF of the larval habitat capture point mixel was modeled as the limit of its directional reflectance factor using:

$$R(i,v):\ R(i,v)=\frac{\iint_A R(s)\langle i,s\rangle\langle v,s\rangle I_i(s)I_v(s)ds}{A\cos\theta_i\theta_v} \quad (7)$$

Where ds was a small Lambertian surface element over area A of the QuickBirdmixel; $R(s)$ was the reflectance of ds; i, v, and s represented the directions of illumination and viewing based on the Precambrian rock surface and ripple water reflectance components, respectively. In the model $\langle .,. \rangle$ was the cosine of the phase angle between two directions; θ was the zenith angle of a direction; $I_i(s)$ and $I_v(s)$ were indicator functions, equal to one when ds was illuminated (I_i) or viewed (I_v) or zero otherwise. If a surface exhibits Lambertian reflectance, light falling on it is scattered such that the apparent brightness of the surface to an observer is the same regardless of the observer's angle of view, thus, the surface luminance is isotropic (Schowengerdt, 1997).

Lambert's cosine law states that the radiant intensity or luminous intensity observed from an ideal diffusely reflecting surface or ideal diffuse radiator is directly proportional to the cosine of the angle θ between the observer's line of sight and the surface normal (Pedrotti and Pedrotti, 1993). In this research the luminous intensity of the geo-refernced *S. damnosum s.l.* riverine larval habitat model endmember point varied by direction. We then defined with peak luminous intensity in the normal direction using the cosine law. As the Lambertian assumption held, we then calculated the total luminous flux, F_{tot}, from the peak luminous intensity by integrating the cosine law:

$$F_{tot}=\int_0^{\pi/2}\int_0^{2\pi}\cos(\theta)I_{max}\sin(\theta)\,\mathrm{d}\phi\,\mathrm{d}\theta=2\pi\cdot I_{max}\int_0^{\pi/2}\cos(\theta)\sin(\theta)\,\mathrm{d}\theta=2\pi\cdot I_{max}\int_0^{\pi/2}\frac{\sin(2\theta)}{2}\,\mathrm{d}\theta$$

and so $F_{tot}=\pi\,\mathrm{sr}\cdot I_{max}$ where $\sin(\theta)$ was the determinant of the Jacobian matrix for the unit sphere, and I_{max} is was the luminous flux per steradian. In vector calculus, the Jacobian matrix is the matrix of all first-order partial derivatives of a vector- or scalar-valued function with respect to another vector (Cressie, 1993). Similarly, the peak intensity was $1/(\pi\,\mathrm{sr})$ of the total radiated luminous flux. For quantitation of the Lambertian surfaces, the same factor of $\pi\,\mathrm{sr}$ related the larval habitat luminance to luminous emittance, radiant intensity to radiant flux, and radiance to radiant emittance. Solvingour double integral equation revealed s that ds was integrated over the decomposed QuickBird mixel (that is, the footprint of the sensor' sinstantaneous field of view (iFOV)) .

In this research, there were two kinds of prominent habitat surfaces in the sub-mixel spectra; A-background surface (that is, Precambrian rock) and surface ripple water-which were represented by Lambertian reflectance G and C, respectively. We then re-wrote equation (7) as:

$$R(i,v)=K_gG+\frac{C}{A}\iint_{A_c}\frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v}ds,$$

where $K_g=A_g/A$ which was the proportion of background spectral data illuminated and viewed by the georefernced, QuickBird, imaged, *S. damnosum s.l.,* riverine, larval, habitat, capture points attributes. In this equation the union of A_g and A_c were the intersection of the dataset of the larval, riverine, habitat, capture, point, surface elements which were illuminated and viewed, only when v and i coincided. The directional reflectance of the habitat scene depended also on the Precambrian rock and ripple water reflectance values related to G and C.

In our analyses we focused on the two terms of $R(i,v)=K_gG+\frac{C}{A}\iint_{A_c}\frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v}ds$. The first term described how the sunlit background proportion proceeded to a maximum point as viewing and illumination positions in the hemisphere coincided. The second term in the model described how the sunlit *S. damnosum s.l.*riverine larval habitat capture point surface composed of the Lambertian facets, became maximally exposed to view at the hotspot, while those facets on tops became dominant at large viewing zenith angles. The hot spot correlation effect refers to the observed brightening which can occur when viewing a scene from the same direction as the solar illumination (Burrough and McDonnell, 1998) which for robust, predictive, spatiotemporal, arthropod-related, infectious disease modeling is commonly noted in the visible and NIR spectral regions (Jacob et al., 2011a). We then analyzed how the first term K_gG varied with illumination and viewing geometry. As in Strahler and Jupp (1990), we assumed that the spatial object of interest (that is, *S. damnosum s.l.* habitat) and its associated georeferenced, explanatory, spectral, predictor, covariate coefficient had the shape of a spheroid, with vertical half-axis equal to b, horizontal radius equal to R, and a height to the center of the spheroid h. To accommodate the spheroidal shape in the derivations of the shadowed,riverine, larval,habitat areas, we used the transformation $\theta'=\tan^{-1}\left(\frac{b}{R}\tan\theta\right)$. We solved this equation by replacing θ with the angle that would generate the same shadow area for a sphere. For simplicity, we assumed that the centers of the spheroids were randomly distributed in depth from h_1 to h_2 over A. We then assumed that G and C were constants and also they were as average signatures over A_g and A_c for properly modelling K_g and $K_c=A_c/A$. Next, the equation $R(i,v)=K_gG+\frac{C}{A}\iint_{A_c}\frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v}ds$ was employed

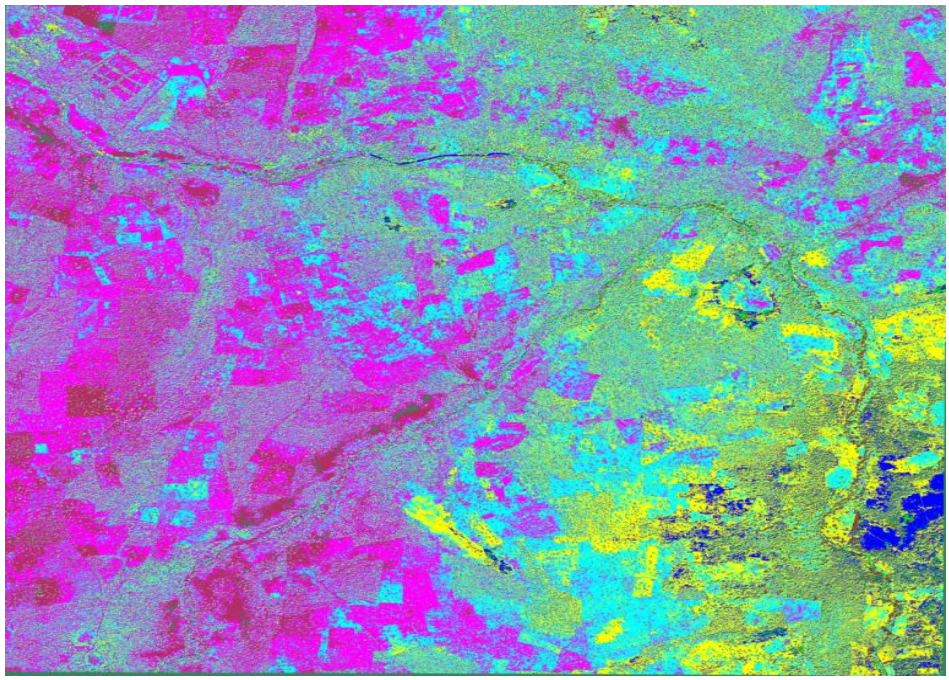


Figure 6. Object-based ENVI classification of the Dienkoa-Chutes riverine breeding study site.

.

whereby K_g was expressed in a Boolean model and $K_g = e^{-\lambda\pi R^2\left[\sec\theta_i' + \sec\theta_v' - \overline{O}(\theta_i,\theta_v,\phi)\right]}$ where $\overline{O}(\theta_i,\theta_v,\phi,)$ represented the average of the overlap function $O(\theta_i,\theta_v,\phi,h)$ between illumination and viewing shadows of the spatiotemporal, field-sampled, *S. damnosum s.l.,* riverine, larval, habitat, capture point and its associated within-canopy structures (for example, Precambrian rock and spectral ripple water components). Furthermore, ϕ was the difference in azimuth angle between viewing and illumination positions of the QuickBird classified objects associated to the geo-refernced capture point habitat. To simplify the equation, we approximated the overlap function by the overlap area and center positions of the ellipses. This approximation is justified when solar zenith and viewing zenith angles are not too large (Strahler and Jupp, 1990). In the case of long ellipsoidal shadows, however, this approximation could have overestimated the width of the riverine, habitat capture point hotspot in the Azimuthal direction and underestimated the width of the hotspot in the Azimuthal direction. To improve the accuracy and to preserve the proper hotspot width information, we developed another approximation as follows. We used the equations $\phi = 0$ *or* $\phi = \pi$ First, we considered the overlap function in the principal plane. We used W $\phi = 0$ or π as the elliptical illumination and then viewed shadows that were aligned in the same direction. The overlap area was approximated by an ellipse with one axis equal to the overlap length and the other equal to the georeferenced, *S. damnosum s.l.,*riverine, larval, habitat, width, encompassing the Precambrian rock and ripple, water, mixel, spectral components which yielded

$$O(\theta_i,\theta_v,\phi) = \frac{1}{2}\left[\sec\theta_i + \sec\theta_v - \frac{h}{b}\left|\tan\theta_i - \tan\theta_v\cos\phi\right|\right]$$

. By so doing, we were able to determine the waveband spectral signature for the riverine larval habitat based on a scattergram (Figure 6).

Interpolation analyses

Multiple, spatial, explanatory, linearized predictors were then generated from the *S. damnosum s.l* .riverine, larval, habitat, capture point, spatiotemporal-endmembers and its associated Precambrian rock and rippled water spectrally decomposed components using an ordinary kriged-based interpolator. The algorithms for our interpolation have already been described in

Jennsen (2005). Briefly, in this research, the dependent variable was the residualized, spectral, emissivity, estimates, rendered from the decomposition of the QuickBird mixel which was transformed to fulfill the diagnostic normality test prior for performing the kriging. Ordinary kriging was selected to interpolate the value $Z(x_0)$, an *S. damnosum s.l.* riverine larval habitat capture point canopy and its associated Precambrian rock and spectral ripple water components,$Z(x)$, at an unobserved, habitat, location x_0 from the field and remote-sampled, explanatory, predictor, covariate, coefficient estimates and $z_i = Z(x_i)$, i = 1..., *n* at nearby habitat locations, x_1, xn. In this research, ordinary kriging was computed as a linear unbiased estimator, $\check{Z}(x_o)$ 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. The kriging estimator was given by a linear combination of the algorithm:

$$\hat{Z}(x_o) = \sum_{i=1}^{n} w_i(x_o)Z(x_i)$$

employing the spectral, riverine, larval, habitat, endmember dataset of $z_i = Z(x_i)$ with weights $w_i(x_o)$, i = 1,...,*n* chosen, such that the variance in the spectral autoregressive model was calculated using:

$$\sigma_k^2(x_0) := Var(\hat{Z}(x_0) - Z(x)) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i(x_0)w_j(x_0)c(x_i,x_j) + Var(Z(x)) - 2\sum_{i=1}^{n} w_i(x_0)c(x_i,x_0)$$

which was further minimized using:

$$E[\hat{Z}(x) - Z(x)] = \sum_{i=1}^{n} w_i(x_0)\mu(x_i) - \mu(x_0) = 0$$

.

Spatial analyses

An autoregressive (p) model was then constructed in SAS/GIS, thereafter using $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$, where $\varphi_1, \cdots, \varphi_p$ were the georeferenced study site geosampled parameter estimators, c was a constant, and the random variable and ε_t was white noise. An AR is essentially an all-pole infinite impulse response filter with some additional interpretation placed on it (Griffith, 2003). In this research, some constraints were necessary on the values of the parameter estimators of the model in order that the residuals remained stationary. For example, processes in the autoregressive models generated at each 5 km interval from the capture point with $|\varphi_1| \geq 1$ were not stationary. The notation MA (q) was then also constructed to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where the $\theta_1, \ldots, \theta_q$ were the riverine estimators, μ was the expectation of X_t (often assumed to equal 0), and the ε_t, ε_{t-1},...were white noise error terms. In this research the notation ARMA (p, q) referred to the model with p autoregressive terms and q moving-average terms. This model contained the AR (p) and MA (q) models which was expressed as:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \vartheta_i \varepsilon_{t-i}.$$

In this research, the error terms ε_t were assumed to be independent identically distributed (i.i.d.), randomized, variables, sampled from a normal distribution with zero mean: $\varepsilon_t \sim N(0,\sigma^2)$. The spatially-dependent, transmission-oriented, endemic, models were then specified in terms of the lag operator L. In these terms then the AR (p) models was provided by:

$$\varepsilon_t = \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \varphi X_t$$ where φ represented the polynomial $\varphi = 1 - \sum_{i=1}^{p} \varphi_i L^i$.

The MA(q) model was then given by the equation $X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t = \theta \varepsilon_t$, where θ represented the polynomials. Finally, the combined ARMA (p, q) endemic models were provided by:

$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$, or more concisely, $\varphi X_t = \theta \varepsilon_t$.

Thereafter, a misspecification perspective for the estimation, endemic, transmission-oriented, risk based, distribution model was constructed in SAS/GIS using $y = X\beta + \varepsilon^*$ (that is, regression equation) for decomposing ε^*, into a white-noise component, ε, (that is, autocorrelation-oriented disturbances) and a set of unspecified and/or misspecified sub- models that had the structure $y = XB + \underbrace{E\gamma + \varepsilon}_{=\varepsilon^*}$. White noise in a spatiotemporal-sampled, arthropod-related, infectious disease, larval, habitat, epidemiological, risk model is a univariate or multivariate discrete-time stochastic process whose terms are i.i.d. with a zero mean (Jensen, 2003).

RESULTS

Initially, we constructed a Poisson regression models using the spatiotemporal-sampled, district-level, covariate, coefficient, measurement values to determine covariate coefficients of significance with each ArcGIS classified Euclidean-distance dependent zone. Our model was generalized by introducing an unobserved heterogeneity term for each spatiotemporal, field-sampled, *S. damnosum s.l.* related, riverine, larval, habitat observation i. The weights were assumed to differ randomly in a manner that was not fully accounted for by the other time series-dependent covariate coefficients. In this research this process was formulated as $E(y_i|\mathbf{x}_i, \tau_i) = \mu_i\tau_i = e^{\mathbf{x}_i\beta + \varepsilon_i}$ where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ was independent of the vector of regressors $\mathbf{x}_i$ for each ArcGIS delineated onchocerciasis, endemic, transmission zone. Then the distribution of y_i was conditional on $\mathbf{x}_i$ and had a Poisson specification with conditional mean where the conditional variance was:

$\mu_i\tau_i$: $f(y_i|\mathbf{x}_i, \tau_i) = \dfrac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$.

We then let $g(\tau_i)$ be the probability density function of τ_i. Then, the distribution $f(y_i|\mathbf{x}_i)$ was no longer conditional on τ_i. Instead it was obtained by integrating $f(y_i|\mathbf{x}_i, \tau_i)$

with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i$. We then found that an analytical solution to this integral existed in the Poisson-relatedendemic model when τ_i was assumed to follow a gamma distribution. The models also revealed that γ_i, was the vector of the sampled covariate coefficients while $\mathbf{x}_i$, was independently Poisson distributed with:

$$P(Y_i = y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \ldots$$

and the mean parameter, that is, the mean number of sampling events per spatiotemporal period was given by $\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ was a $(k+1) \times 1$ parameter vector. The intercept in the model was then β_0 and the coefficients for the k regressors were $\beta_1, \ldots, \beta_k$. Taking the exponential of $\mathbf{x}_i'\boldsymbol{\beta}$ ensured that the mean parameter μ_i was non-negative. Thereafter, the conditional mean was provided by $E(y_i|\mathbf{x}_i) = \mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. The parameter estimators were then evaluated using $\ln[E(y_i|\mathbf{x}_i)] = \ln(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$. Note, in this research, that the conditional variance of the count random variable was equal to the conditional mean (that is, equidispersion) ($V(y_i|\mathbf{x}_i) = E(y_i|\mathbf{x}_i) = \mu_i$). In a log-linear model the logarithm of the conditional mean is linear (Hosmer et al., 2002). The marginal effect of any district-level, explanatory regressor in the models was then provided by:

$$\frac{\partial E(y_i|\mathbf{x}_i)}{\partial x_{ji}} = \exp(\mathbf{x}_i'\boldsymbol{\beta})\beta_j = E(y_i|\mathbf{x}_i)\beta_j$$

In the model we noticed that a one-unit change in the jth regressor led to a proportional change in the conditional mean $E(y_i|\mathbf{x}_i)$ of β_j

Further, we found that given the Poisson process in our endemic, transmission-oriented, risk-based model, the limit of the binomial distribution in the sampled parameter estimators was:

$$P_p(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

Viewing the distribution as a function of the expected number of successes [$v = Np$] instead of the sample size N for fixed p, transformed the equation to:

$$P_{v|N}(n|N) = \frac{N!}{n!(N-n)!} \left(\frac{v}{N}\right)^n \left(1 - \frac{v}{N}\right)^{N-n}$$

As the sample size N become larger, the distribution then approached P when:

$$\lim_{N\to\infty} P_p(n|N), \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!} \frac{v^n}{N^n} \left(1 - \frac{v}{N}\right)^N \left(1 - \frac{v}{N}\right)^{-n}, \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n} \frac{v^n}{n!} \left(1 - \frac{v}{N}\right)^N \left(1 - \frac{v}{N}\right)^{-n},$$

$$1 \cdot \frac{v^n}{n!} \cdot e^{-v} \text{ and } \frac{v^n e^{-v}}{n!},$$

Note that the sample size N had completely dropped out of the probability function, which had the same functional form for all values of v. Next, the moment-generating function of the Poisson distribution was provided by:

M = $e^{-v} e^{ve^t} = e^{v(e^t-1)}$, M=$v e^t e^{v(e^t-1)}$ and M=$(v e^t)^2 e^{v(e^t-1)} + v e^t e^{v(e^t-1)}$, when R=$v(e^t - 1)$, R'=$v e^t$ so R=$R'(0) = v$.

The raw moments were also computed directly by summation, which yielded an unexpected connection with the exponential polynomial $\phi_n(x)$ and the stirling numbers of the second kind

$$\phi_n(x) = \sum_{k=0}^{\infty} \frac{e^{-x} x^k}{k!} k^n = \sum_{k=1}^{n} x^k S(n, k)$$

which in this research was represented as Dobiński's formula for the Bell polynomial and Bell numbers. This generalized formula revealed $B_n(x) = e^{-x}\sum_{k=0}^{\infty}\frac{k^n}{k!}x^k$, where $B_n(x)$ was a Bell polynomial .Setting $x = 1$gives the special case of the nth Bell number, $B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}$. The formula was derived by dividing the generating function formula for a Stirling number of the second kind $S(n,k)$ and by quantitating by $m!$, yielding $\frac{m^n}{m!} = \sum_{k=1}^{n}\frac{S(n,k)}{(m-k)!}$. Then:

$\sum_{m=1}^{\infty}\frac{m^n}{m!}\lambda^m = \left(\sum_{k=1}^{n}S(n,k)\lambda^k\right)\left(\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}\right)$, and $\sum_{k=1}^{n}S(n,k)\lambda^k = e^{-\lambda}\sum_{m=1}^{\infty}\frac{m^n}{m!}\lambda^m$. (Roman, 1984).

Then in the endemic Euclidean distance–based models: $\nu(1+\nu)$, $\nu\left(1+3\nu+\nu^2\right)$ and $\nu\left(1+7\nu+6\nu^2+\nu^3\right)$.

We then tested the model for over-dispersion with a likelihood ratio test based on the linear distributions. This test quantitated the equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeded the mean. For the negative binomial distribution, the variance = mean + k mean^2 k > = 0 and the negative binomial distribution reduces to Poisson when k = 0 (Haight, 1967). The probability mass function of the negative binomial distributions with a gamma distributed mean was then expressed as:

$$f(k) = \Pr(X = k) = \binom{k+r-1}{k}(1-p)^r p^k \quad \text{for } k = 0, 1, 2, \ldots$$

In this equation, the quantity in parentheses was the binomial coefficient, and was equal to:

$$\binom{k+r-1}{k} = \frac{(k+r-1)!}{k!\,(r-1)!} = \frac{(k+r-1)(k+r-2)\cdots(r)}{k!}.$$

In this research, this quantity was also alternatively written as:

$$\frac{(k+r-1)\cdots(r)}{k!} = (-1)^k\frac{(-r)(-r-1)(-r-2)\cdots(-r-k+1)}{k!} = (-1)^k\binom{-r}{k}$$

for explaining “negative binomialness’ in the onchocerciasis, endemic, transmission-oriented, risk model. Results from both a Poisson and a negative binomial (that is, a Poisson random variable with a gamma distrusted mean) revealed that the explanatory, predictor, covariate coefficients were highly significant, but furnished virtually no predictive power. In other words, the sizes of the population denominators were not sufficient to result in statistically significant relationships, while the detected relationships were inconsequential. For χ^2 fitting, the likelihood wa sprovided by:

$L = \prod\left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2}\exp\left(-\sum\frac{(y_i - f(\mathbf{x}))}{2\sigma_i^2}\right)$ (that is, $\ln L - \ln\left(\prod\left(\frac{1}{2\pi\sigma_i}\right)^{1/2}\right) - \frac{1}{2}\sum\frac{(y_i - f(\mathbf{x})}{\sigma_i^2} = C - \chi^2/2$)

where C wasa constant independent of the model and dependent only on the use of particular data points (that is, it does not change if the data do not change). In this research we employed a correlation analysis method to check the cross correlation between the input and output signals as an estimation of the impulse response, as shown by the following equation:

$$y(k) = \sum_{n=0}^{\infty}u(k-n)h(n) + e(k)$$

The input signal must be zero-mean white noise with a spectral density that is equally distributed across the whole frequency range (Cressie 1993). The SI estimate impulse response VI can prewhiten input signals that are not white noise (Hosmer et al., 2000). Thus, assuming the input $u(k)$ of the system was stochastic process and statistically independent of the disturbance $e(k)$, we assumed the following equation was true:

$$R_{uy}(\tau) = \sum_{k=0}^{\infty} R_{uu}(k-\tau)h(k)$$

When R_{uy} represented the cross-correlation function between the stimulus signal $u(k)$ and the response signal $y(k)$, as defined by:

$$R_{uy}(\tau) = \frac{1}{N} \sum_{k=min(\tau,0)}^{N-max(\tau,0)-1} y(k+\tau)u(k)$$

When R_{uu} represented the autocorrelation of the stimulus signal $u(k)$, as defined by the following equation:

$$R_{uu}(\tau) = \frac{1}{N} \sum_{k=0}^{N-\tau-1} u(k+\tau)u(k)$$

We employed N as the number of sampled, onchocerciasis-related, endemic, transmission-oriented, risk-based, georeferenced data points. If the stimulus signal is a zero-mean white noise signal, the autocorrelation function reduces to the following equation. $R_{uu}(\tau) = \sigma_u^2 \delta(\tau)$ where σ_u is the standard deviation of the stimulus white noise and $\delta(\tau)$ is the Dirac function (Cressie, 2993). Thereafter, we substituted $R_{uu}(\tau)$ into the cross-correlation function between the stimulus signal $u(k)$ and the response signal $y(k)$ which yielded the following equation:

$$R_{uy}(\tau) = \sigma_u^2 \sum_{k-0}^{\infty} \delta(k-\tau)h(k) = \sigma_u^2 h(\tau)$$

We rearranged the terms of this equation to obtain the following equation defining the impulse response:

$$h(k) = \frac{R_{uy}(k)}{\sigma_u^2}$$

$h(k)$.

The correlation analysis method then estimated the impulse response to be robust but only when the input signal $u(k)$ was a zero-mean white noise signal. However, the input signal is not white noise in most real-world applications (Cressie, 1993). Therefore, we preconditioned the input $u(k)$ and output $y(k)$ signals before we applied them to our e correlation analysis method. We then generated a set of $k+1$ transmission-oriented risk based data points $(x_0, y_0), \ldots, (x_k, y_k)$ where no two x_j were the same employing an interpolation polynomial in the Newton form. By so doing, a linear combination of Newton basis polynomials:

$$N(x) := \sum_{j=0}^{k} a_j n_j(x)$$

[That is,] with the Newton basis polynomials was defined as $n_j(x) := \prod_{i=0}^{j-1}(x - x_i)$ for $j > 0$ and $n_0(x) = 1$. The coefficients were then defined as $a_j := [y_0, \ldots, y_j]$ where $[y_0, \ldots, y_j]$ which in this research was represented using the notation for divided differences. As such, the endemic, transmission-oriented, Newton polynomial was then written as

$$N(x) = [y_0] + [y_0, y_1](x - x_0) + \cdots + [y_0, \ldots, y_k](x - x_0)(x - x_1)\cdots(x - x_{k-1}).$$

The Newton polynomial was then expressed in a simplified form when $x_0, x_1, \ldots, x_k$ which in this research was

arranged consecutively with equal spacing. Introducing the notation $h = x_{i+1} - x_i$ for each $i = 0, 1, \ldots, k-1$ and $x = x_0 + sh$, then rendered the difference $x - x_i$ which was then re-written as $(s-i)h$. So the onchocerciasis, endemic, transmission-oriented, Newton polynomial became:

$$N(x) = [y_0] + [y_0, y_1]sh + \cdots + [y_0, \ldots, y_k]s(s-1)\cdots(s-k+1)h^k$$ which in this research was:

$$= \sum_{i=0}^{k} s(s-1)\cdots(s-i+1)h^i[y_0, \ldots, y_i] = \sum_{i=0}^{k} \binom{s}{i} i!h^i[y_0, \ldots, y_i] N(x) = \sum_{i=0}^{k} \binom{s}{i} i!h^i[y_0, \ldots, y_i]$$

(that is, Newton forward divided difference formula). The polynomial interpolation was then used to construct the polynomial of degree $\leq n$ that passes through the *n*+1 sampled points $(x_k, y_k) = (x_k, f(x_k))$, for $k = 0, 1, \cdots, n$. If multiple "centers" $x_0, x_1, \ldots, x_n$ are used, then the result is the so called Newton polynomial (Hosmer et al., 2000). We then assumed that $f \in C^{n+1}[a, b]$ and $x_k \in [a, b]$ for $k = 0, 1, \cdots, n$ were distinct spatiotemporal-sampled values. Then $f(x) = P_n(x) + R_n(x)$ where $P_n(x)$ was a polynomial which in this research was used to approximate $f(x)$ and also:

$$P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) + a_n(x - x_0)(x - x_1)(x - x_2) \ldots (x - x_{n-1}).$$

Thereafter, we wrote $f(x) \approx P_n(x)$. Our model revealed that the Newton polynomial went through the $n+1$ onchocerciasis, transmission–oriented, risk-based, sampled epidemiological points $\{(x_k, y_k)\}_{k=0}^{n}$, (that is, $P_n(x_k) = f(x_k)$ for $k = 0, 1, \ldots, n$). The remainder term $R_n(x)$ had the form:

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)(x - x_1)(x - x_2) \ldots (x - x_{n-1})(x - x_n)$$, for any model parameter estimator value when $c = c(x)$ such that lay in the interval $[a, b]$. The covariate coefficients a_i were then constructed using divided differences.

In this research, the divided differences for a function $f[x]$ in our onchocerciasis, endemic, transmission-oriented, risk-based, landscape, distribution, epidemiological model were defined as:

$$f[x_{i-1}, x_i] = \frac{f[x_i] - f[x_{i-1}]}{x_i - x_{i-1}}, \quad f[x_{i-2}, x_{i-1}, x_i] = \frac{f[x_{i-1}, x_i] - f[x_{i-2}, x_{i-1}]}{x_i - x_{i-2}},$$

$$f[x_{i-3}, x_{i-2}, x_{i-1}, x_i] = \frac{f[x_{i-2}, x_{i-1}, x_i] - f[x_{i-3}, x_{i-2}, x_{i-1}]}{x_i - x_{i-3}}$$

and also

$$f[x_{i-j}, x_{i-j+1}, \ldots, x_i] = \frac{f[x_{i-j+1}, \ldots, x_i] - f[x_{i-j}, \ldots, x_{i-1}]}{x_i - x_{i-j}}$$

The divided difference formulae were then used to construct the endemic, transmission-oriented, risk-related, epidemiological model divided difference table as:

$x_i\ f[x_i]\ f[x_{i-1}, x_i]\ f[x_{i-2}, x_{i-1}, x_i]\ f[x_{i-3}, x_{i-2}, x_{i-1}, x_i]\ f[x_{i-4}, x_{i-3}, x_{i-2}, x_{i-1}, x_i]\ ; x_0\ f[x_0]\ , x_1\ f[x_1]\ , x_2\ f[x_2]\ f[x_0, x_1]\ , x_3\ f[x_3]\ f[x_1, x_2]\ , x_4\ f[x_4]\ f[x_2, x_3]\ f[x_0, x_1, x_2]\ , f[x_3, x_4]\ f[x_1, x_2, x_3]\ f[x_2, x_3, x_4]\ f[x_0, x_1, x_2, x_3]\ f[x_1, x_2, x_3, x_4]\ f[x_0, x_1, x_2, x_3, x_4]\ , x_i\ f[x_i]\ f[x_{i-1}, x_i]\ f[x_{i-2}, x_{i-1}, x_i]\ f[x_{i-3}, x_{i-2}, x_{i-1}, x_i]\ f[x_{i-4}, x_{i-3}, x_{i-2}, x_{i-1}, x_i]$

The coefficient of our endemic, transmission-oriented, risk-based, landscape, distribution model was then directly related to the Newton polynomial when $P_n(x)$ was $a_i = f[x_0, x_1, \ldots, x_i]$. The coefficient also was the top element in the column of the *i*-th divided differences. The Newton polynomial of degree ≤ n then passed through n+1, risk-based, sampled points (that is, $(x_k, y_k) = (x_k, f(x_k))$, for $k = 0, 1, \cdots, n$) which was then quantitated as:

$$P_n(X) = a_0 + a_1 (X - X_0) + a_2 (X - X_0)(X - X_1) + a_3 (X - X_0)(X - X_1)(X - X_2)$$

$$+ a_n (x - x_0)(x - x_1)(x - x_2) \ldots (x - x_{n-1}) .$$

The form the Newton polynomials of degree for the function $f[x] = \mathrm{Cos}[x]$ over the interval $[x_0, x_n]$ was then used along with equally spaced nodes selected from the following list:

$$\{(x_k, y_k)_{k=0}^{5}\} = \left\{\{0, 1\}, \left\{\frac{1}{5}, \mathrm{Cos}\left[\frac{1}{5}\right]\right\}, \left\{\frac{2}{5}, \mathrm{Cos}\left[\frac{2}{5}\right]\right\}, \left\{\frac{3}{5}, \mathrm{Cos}\left[\frac{3}{5}\right]\right\}, \left\{\frac{4}{5}, \mathrm{Cos}\left[\frac{4}{5}\right]\right\}, \{1, \mathrm{Cos}[1]\}\right\}$$ and so on. Thereafter, we let:

$$\pi_n(x) = \prod_{k=0}^{n} (x - x_k), \qquad f(x) = f_0 + \sum_{k=1}^{n} \pi_{k-1}(x)\,[x_0, x_1, \ldots, x_k] + R_n,$$

and as such: where $[x_1, \ldots]$ was a divided difference, and the remainder was: $R_n(x) = \pi_n(x)\,[x_0, \ldots, x_n, x] = \pi_n(x)\,\dfrac{f^{(n+1)}(\xi)}{(n+1)!}$ for $x_0 < \xi < x_n$ (Figure 7).

The error terms corresponding to these onchocerciasis, endemic, transmission-oriented, explanatory, risk-based landscape model had the following useful bounds on their magnitude:

(i). $|R_1(x)| \le \dfrac{M_2}{8} h^2$ which was valid for $x \in [x_0, x_1]$,

(ii). $|R_2(x)| \le \dfrac{M_3}{9\sqrt{3}} h^3$ which was valid for $x \in [x_0, x_2]$,

(iii). $|R_3(x)| \le \dfrac{M_4}{24} h^4$ which was valid for $x \in [x_0, x_3]$,

(iv). $|R_4(x)| \le \dfrac{\sqrt{4750 + 290\sqrt{145}}}{3000} M_5 h^5$ which was valid for $x \in [x_0, x_4]$,

(v). $|R_5(x)| \le \dfrac{10 + 7\sqrt{7}}{1215} M_6 h^6$ which was valid for $x \in [x_0, x_5]$

The seasonal-sampled onchocerciasis related-endemic, transmission-oriented, polynomials of Newton basis e_k were then defined by:

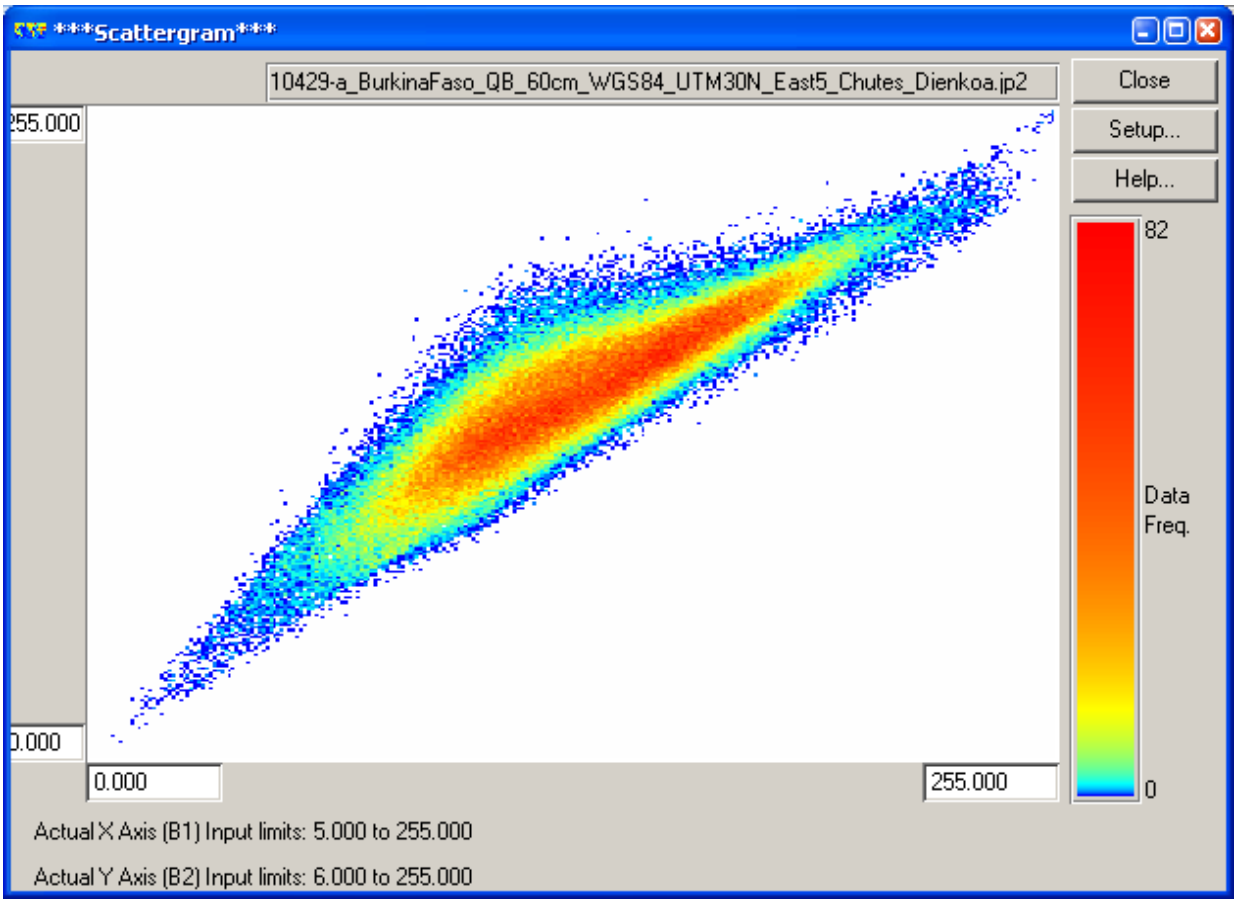


Figure 7. An endmemberspectral signature of the georeferenced *S. damnosum s.l.* riverine habitat capture point.

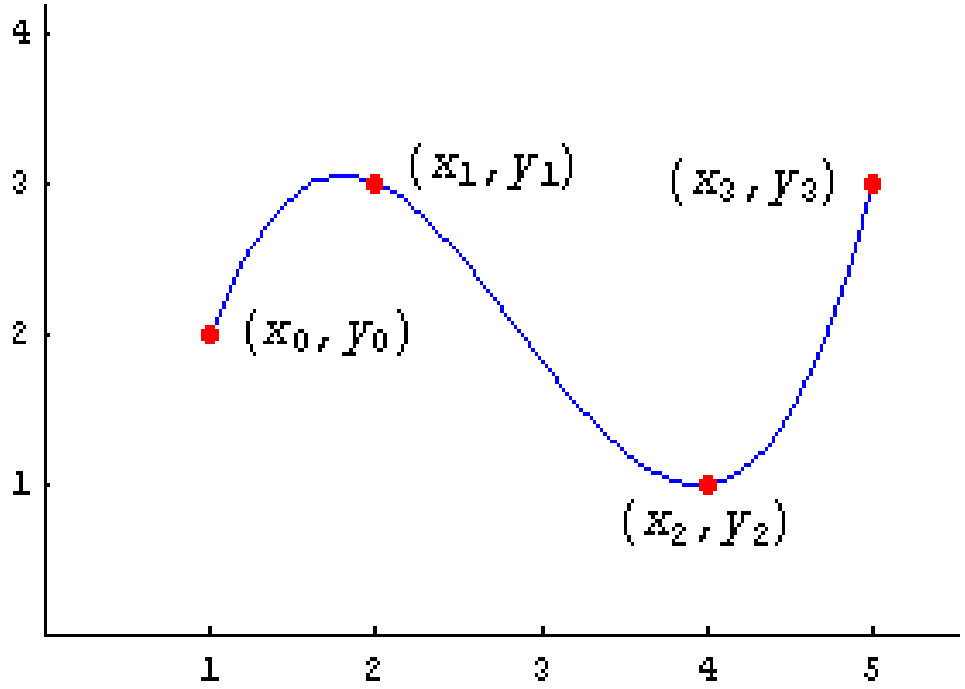


Figure 8. A cubic curve below illustrating the risk-based Newton polynomial of degree *n*=3.

$$e_k(x) = \prod_{i=0}^{k-1}(x - x_i) = (x - x_0)(x - x_1)\cdots(x - x_{k-1}), \quad k = 1, \ldots, n.$$

employing the following convention: $e_0 = 1$. Moreover

$$\begin{array}{lllllll} e_1 & = & (x - x_0) & e_2 & = & (x - x_0)(x - x_1) & e_3 & = & (x - x_0)(x - x_1)(x - x_2) \\ e_n & = & (x - x_0)(x - x_1)\cdots(x - x_{n-1}) & & & & & & \end{array}$$

.

The set of robust transmission-oriented, risk-based, epidemiological polynomials [that is, $(e_k)_{0\le k\le n}$] were the basis of P_N , which in this research represented the space of polynomial of degree and was at most equal to *n*. Indeed, they constituted an echelon-degree set of $(n+1)$ polynomial. Newton interpolation of degree n in this research was thereafter related to the subdivision

$\{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$ = $P_n(x) = \sum_{k=0}^{n} a_k e_k(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1)$ where

$P_n(x_i) = f(x_i), \quad \forall i = 0, \ldots, n.$

We then determined the explanatory, predictor, covariate coefficients $(a_k)_{0\le k\le n}$ using divided differences. Newton's interpolation polynomial of degree n , $P_n(x)$, were then evaluated at x_0 rendering:

$$P_n(x_0) = \sum_{k=0}^{n} a_k e_k(x_0) = a_0 = f(x_0) = f[x_0]$$

We then re-wrote $f[x_i] = f(x_i), \quad \forall i = 0, \ldots, n\ f[x_0]$ as a zero-order divided difference. The onchocerciasis –related, Newton's interpolation, polynomial of degree n , $P_n(x)$ was then evaluated at x_1 rendering:

$$P_n(x_1) = \sum_{k=0}^{n} a_k e_k(x_1) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) \ \ldots + a_n(x - x_0)(x - x_1)\cdots(x - x_{n-1})$$

Hence: $a_1 = \dfrac{f[x_1] - f[x_0]}{x_1 - x_0} = f[x_0, x_1]$.

In our model $f[x_0, x_1]$ was the first -order divided difference. The interpolation polynomial of degree *nPn*(*x*)was then evaluated at x_2 rendering:

$$\begin{array}{l} P_n(x_2) = \sum_{k=0}^{n} a_k e_k(x_2) = a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) = a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) \\ = f[x_0] + f[x_0, x_1](x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) = f[x_2] \end{array}$$

Then: $\alpha_2(x_2 - x_0)(x_2 - x_1) = f[x_2] - f[x_0] - f[x_0, x_1](x_2 - x_0)$

$$\alpha_2 = \frac{f[x_2] - f[x_0]}{(x_2 - x_0)(x_2 - x_1)} - \frac{f[x_0, x_1]}{x_2 - x_1}$$

$$\alpha_2 = \frac{f[x_2] - f[x_0]}{(x_2 - x_0)(x_2 - x_1)} - \frac{f[x_0, x_1]}{x_2 - x_1} \quad \text{and} \quad \alpha_2 = \frac{f[x_0, x_2] - f[x_0, x_1]}{x_2 - x_1}$$

The final following form of the polynomialized,onchocerciasis-related, regression-based,parameter estimators was then:

$$\begin{aligned}
\alpha_2(x_2 - x_0)(x_2 - x_1) &\equiv f[x_2] - f[x_0] \\
\alpha_2(x_2 - x_0)(x_2 - x_1) &\equiv f[x_2] - f[x_0] \\
\alpha_2(x_2 - x_0)(x_2 - x_1) &\equiv f[x_2] - f[x_1] \\
\alpha_2(x_2 - x_0)(x_2 - x_1) &\equiv f[x_2] - f[x_1] \\
\alpha_2(x_2 - x_0)(x_2 - x_1) &\equiv f[x_2] - f[x_1] \\
\alpha_2(x_2 - x_0) &\equiv \frac{f[x_2] - f[x_1]}{x_2 - x_1} \\
\alpha_2(x_2 - x_0) &\equiv f[x_1, x_2] - f[
\end{aligned}$$

$$\alpha_2 = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = f[x_0, x_1, x_2]$$

Hence, in this research, was the second-order divided difference. By recurrence, we then obtained:

$$\alpha_k = \frac{f[x_1, \ldots, x_k] - f[x_0, \ldots, x_{k-1}]}{x_k - x_0} = f[x_0, \ldots, x_k]$$

which was then the k[th]-order divided difference. The Newton's interpolation polynomial of degree was then obtained via the successive divided differences:

$$P_n(x) = f[x_0] + \sum_{k=1}^{n} f[x_0, \ldots, x_k] e_k(x)$$

.

Thereafter, for quantizing interpolation error in our model, we assumed that:

$f \in \mathcal{C}^n([a, b])$ and $x \in [a, b]$.

We then let *I* be the closed set defined by:

$I = [\min(x, x_0), \max(x, x_n)]$ (that is, the smallest closed set containing x and the x_i's). We employed:

$\exists \xi \in I / f[x_0, \ldots, x_n] = \frac{f^n(\xi)}{n!}$ then we let: $d(x) = f(x) - p(x).$

The quantized, interpolation, error factor and its complement were then:

$\eta_q = \eta + \nu$ and $\overline{\eta}_q = \overline{\eta} - \nu$ where, $\eta, \overline{\eta}$ were unsigned; {that is, $|\nu| \leq 2^{-(n_\eta+1)}$].

The interpolated coefficient then rendered $\hat{h}_{qq}(t) = (\overline{\eta} - \nu)[h(t_0) + \epsilon_0] + (\eta + \nu)[h(t_1) + \epsilon_1]$ and $= \hat{h}(t) + \overline{\eta}\epsilon_0 + \eta\epsilon_1 + \nu[h(t_1) - h(t_0)]$; when the second-order errors $V\varepsilon_0$ and $V\varepsilon_1$ were dropped. Since $|h(t_1) - h(t_0)| \leq M_1$, we obtained the error bound for the onchocerciasis-related endemic transmission-oriented model using: $|e_{qq}(t)| \leq 2^{-n_c} + 2^{-(n_\eta+1)} M_1 + \frac{1}{8} M_2.$

By successively applying Rolle's theorem (*n*times) in the spatiotemporal, infectious disease model, $d^{(n)}(x)$ equaled zero at any given sampled point $\xi \in I$: $d^{(n)}(\xi) = 0.$ Thus, we had: $f^{(n)}(\xi) = P_n^{(n)}(\xi)$. Since the seasonal-sampled, onchocerciasis, explanatory, covariate coefficients of x^n in P_n in this research was $f[x_0, \ldots, x_n]$, $f^{(n)}(\xi) = P_n^{(n)}(\xi) = n! \cdot f[x_0, \ldots, x_n]$ hence $f[x_0, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!}$.

In calculus, Rolle's theorem essentially states that a differentiable function attains equal values at two distinct points which must have a point somewhere between them where the first derivative (that is, the slope of the tangent line to the graph of the function) is zero (Ross, 2007). We then assumed that: $f \in \mathcal{C}^{n+1}([a,b])$ and $x \in [a,b]$ in the transmission-oriented, risk-based, epidemiological, risk model .We then let *I* be the closed set defined by: $I = [\min(x, x_0), \max(x, x_n)]$ (that is, the smallest closed empirical data set containing *x*) and then solved for: We then let $\hat{x} \in [a,b]$ and then assumed that $\hat{x} \neq x_i$ for performing a Lagrange polynomial interpolation for mapping $\forall x \in [a,b], \quad \exists \xi \in I / f(x) - P_n(x) = \frac{f^{n+1}(\xi)}{(n+1)!} \prod_{i=0}^{n} (x - x_i)$ = the endemic, transmission-oriented regions within the riverine study site. In numerical analysis, Lagrange polynomials are used for polynomial interpolation. For a given set of distinct points x_j and numbers y_j, the Lagrange polynomial is the polynomial of the least degree that at each point x_j assumes the corresponding value y_j (that is, the functions coincide at each point) (Cressie, 1993). We then considered the unique polynomial *Pn+1* of degree (n+1) which interpolated *f* at the transmission-oriented, risk-based, regression points [that is, $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n), (\hat{x}, f(\hat{x}))$] and P_{n+1} which verified:

$$\begin{cases} P_{n+1}(x_i) = f(x_i), & \forall i = 0, \ldots, n \\ P_{n+1}(\hat{x}) = f(\hat{x}). \end{cases}$$

The polynomial P_{n+1} was then written as: $P_{n+1}(x) = P_n(x) + (x - x_0) \ldots (x - x_n) f[x_0, \ldots, x_n, \hat{x}]$

$$f[x_0, \ldots, x_n, \hat{x}] = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

According to our model , therefore, by setting the expressions:

$$x = \hat{x}, \quad P_{n+1}(\hat{x}) = f(\hat{x}), \quad f(\hat{x}) = P_n(\hat{x}) + (\hat{x} - x_0)\ldots(\hat{x} - x_n)\frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

In our analyses we found three 3 data points $\{(0,1),(2,5),(4,17)\}$, which determined the Newton interpolation polynomial of degree 3 which passed through the following points:

$$x_0 = 0 \quad f[x_0] = 1, \begin{bmatrix} x_0 = 0 & f[x_0] = 1 & \\ x_1 = 2 & f[x_1] = 5 & f[x_0, x_1] = \frac{5-1}{} = 2 \end{bmatrix}$$

which represented the:

$$x_2 = 4 \quad f[x_2] = 17 \quad f[x_1, x_2] = \frac{17-5}{4-2} = 6 \quad f[x_0, x_1, x_2] = \frac{6-2}{4-0} = 1$$

0 (that is, capture point) from 0 to 5 km hyperendemic range, 5 to 10 km. Thereafter, a mesoendemic range was measured from 5 to 10 km, 10 to 15 km was hypoendemic and after 15 km no transmission.

To validate the Newton polynomial that passed through the transmission-oriented endemic points we tested $i = 0, 1, \cdots, n$ which in this research was performed where:

$d_{i,0} = y_i$ for $i = 0, 1, \ldots, n$ and $d_{i,j} = \frac{d_{i,j-1} - d_{i-1,j-1}}{x_i - x_{i-j}}$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, i$

Newton polynomials were then created "recursively." employing:

$$P_n(x) = P_{n-1}(x) + d_{n,n}(x - x_0)(x - x_1)(x - x_2)\ldots(x - x_{n-1}).$$

The divided difference $f[x_0, x_1, x_2, \ldots, x_n]$, was then denoted $[x_0, x_1, x_2, \ldots, x_n]$), whereby the transmission-oriented risk-based points, x_0, x_1, ..., x_n of a function $f(x)$ was defined by $f[x_0] = f(x_0)$ and $f[x_0, x_1, \ldots, x_k] = \frac{f[x_0, \ldots, x_{n-1}] - f[x_1, \ldots, x_n]}{x_0 - x_n}$ for $n \geq 1$. The first few differences were quantitated using:

$f[x_0, x_1] = \frac{f_0 - f_1}{x_0 - x_1}$, $f[x_0, x_1, x_2] = \frac{f[x_0, x_1] - f[x_1, x_2]}{x_0 - x_2}$ and then:

$$f[x_0, x_1, \ldots, x_n] \frac{f[x_0, \ldots, x_{n-1}] - f[x_1, \ldots, x_n]}{x_0 - x_n}.$$

Thereafter, we defined:

$\pi_n(x) = (x - x_0)(x - x_1)\cdots(x - x_n)$ and then solved $\pi_n'(x_k) = (x_k - x_0)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)$ which rendered the identity: $f[x_0, x_1, \ldots, x_n] = \sum_{k=0}^{n} \frac{f_k}{\pi_n'(x_k)}$.

All the time series-dependent, riverine, spatial, data, feature, attribute points based on the QuickBird mixel encompassing the Precambrian rock and rippled water components were then examined in *n*-dimensional space. A meaningful endmember spectrum for the vertex was calculated from the radiative transfer model residuals employing the vector Euclidean norm to the subspace as defined by the selected, georeferenced, *S. damnosum s.l.,*riverine, larval, habitat, capture point Precambrian rock and rippled water spectral endmembers. To find these candidate mixel spectra, we constructed a QuickBird endmember dataset using *P possible*, consisting of *r* sub- mixel spectral emissivities that were closest to the vertex. Then we generated an spectral endmember subset using $P_{candidate} = \left\{\vec{p}_{(x_1y_1)}, \vec{p}_{(x_2y_2)}, \ldots, \vec{p}_{(x_ry_r)}\right\} (c \leq r, P_{candidate} \in P_{possible})$ which was selected from the *r* pixels ($P_{possible}$) which was subject to conditions based on: $x_i - x_j \leq t\_pixel$ and $y_i - y_j \leq t\_pixel (i \neq j)$. We used the equation $s\_angle(\vec{p}_{(x_iy_i)}, \vec{p}_{(x_jy_j)}) \leq t\_\theta \; (i \neq j)$ as the spectral angle between the epidemiological capture point, Precambrian rock and ripple water sub-mixel radiance and calculated $s\_angle(\vec{p}_{(x_iy_i)}, \vec{p}_{(x_jy_j)}) = cos^{-1}\left(\frac{\vec{p}_{(x_iy_i)} \cdot \vec{p}_{(x_jy_j)}}{|\vec{p}_{(x_iy_i)}||\vec{p}_{(x_jy_j)}|}\right)$ where *t_θ* was the threshold value for the spectral angle beyond which the endmember spectra were not considered similar. The value of *t_θ* was set at 2.5 degrees. The unmixing algorithm identified $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m$ the brightest, sub-mixel, data, feature attribute (i.e., ripple water pixel components) and the darkest, sub-mixel, data attribute (that is, Precambrian rock). The algorithm then iteratively found the remaining endmembers using orthogonal projections until the number of endmembers defined was obtained.

A number of Precambrian rock, ripple water components and the sampled*S. damnosum s.l.*riverine, larval, habitat, capture point, and the endmembers (*m*) were used to find the spectral angle threshold *t_θ* and the spatial threshold (that is, *t_θ* QuickBird mixel) in the SPA algorithm. The first step was to extract the first endmembers. The vector norms of the QuickBird sub-mixel data determined the largest norm value. Sub-mixel heterogeneity at the simplex vertices were then calculated which revealed the radiance data in the image cube. The capture point first endmember was estimated. We then used an orthogonal projection for extraction of all the other related endmember, sub-mixel, *S. damnosum s.l.,*riverine, larval, habitat, capture point Precambrian rock and ripple water spectral components. All the QuickBird, sub-mixel, spectral data was then fractionally calculated based on the visible and NIR spectrum. We used an endmember matrix $U = [\vec{e}_1, \vec{e}_2, \vec{e}_3]$ which projected the QuickBird sub-mixel data into subspace .We noticed that S_{proj} was orthogonal to the space spanned by *U* as $\vec{p}_{(i,f)\_proj} = O\vec{p}_{(i,f)}$, where $\vec{p}_{(i,f)\_proj}$ and $\vec{p}_{(i,f)}$ were the projected and original mixel vector at the georeferenced capture point location (i, f), respectively.

In this research, O was the projection operator, $O = I - UU^+$where *I* was the identity matrix and U^+ was the pseudo inverse of *U,* which was denoted by $U^+ = (U^T U)^{-1} U^T$? We then validated the endmember matrix (that is, $U = [\vec{e}_1, \vec{e}_2, \vec{e}_3]$). We calculated the change of the simplex volume with each subspace projection. The volume of the simplex was then derived. The volume increase was determined by the spectral contrasts between the endmembers. Here, C_{i-1} and C_i denoted the simplexes defined by the original endmember set, $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_{i-1}, \vec{e}_i\}$ and the ratio of the volumes of C_{i-1} and C_i was calculated as $v\_ratio_i = \frac{V(C_i)}{V(C_{i-1})}$. $(3 \leq i \leq m)$. Each *S. damnosum s.l.* riverine habitat endmember proportion was then calculated.

In our analyses, the BRDF of the decomposed *S. damnosum s.l.* habitat mixel using the geometric-optical model was modeled as the limit of its directional reflectance factor using:

$$R(i,v) \quad R(i,v) = \frac{\int\int_A R(s)\langle i,s\rangle\langle v,s\rangle I_i(s) I_v(s) ds}{A\cos\theta_i\theta_v} \quad (8)$$

Where *ds* was a small Lambertian surface element over area A of the QuickBirdmixel; $R(s)$ was the reflectance of ds; i, v, and s represented the directions of illumination and viewing based on the Precambrian rock surface and rippled water,spectral,reflectance components, respectively. In our model $\langle .,. \rangle$ was the cosine of the phase angle between two directions; θ was the zenith angle of a direction; $I_i(s)$ and $I_v(s)$ were indicator functions, equal to one when ds was illuminated (I_i) or viewed (I_v) or zero otherwise. If a surface exhibits Lambertian reflectance, light falling on it is scattered such that the apparent brightness of the surface to an observer is the same regardless of the observer's angle of view, thus, the surface luminance is isotropic (Schowengerdt, 1997). Solving the double integral equation revealed that ds was integrated over

the decomposed QuickBird mixel [that is, the footprint of the sensor's instantaneous field of vision (iFOV)]. We noticed that there were two kinds of prominent riverine larval habitat surfaces in the sub-mixel, endmember, spectra; A-background, surface (that is, Precambrian rock) and spectral ,surface-oriented, ripple, water, data, feature attributes were represented by Lambertian reflectance G and C, respectively. We then re-wrote equation (8) as:

$$R(i,v) = K_g G + \frac{C}{A}\iint_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds \,, \quad \text{where}$$

$K_g = A_g / A$ was the proportion of background spectral data illuminated and viewed rendered by the QuickBird imaged capture point attributes. In this equation the union of *Ag* and A_c were the intersection of the dataset of the capture point surface elements which were illuminated and viewed, only when v and i coincided. The directional reflectance of the habitat scene depended also on the Precambrian rock and ripple water reflectance related to *G* and *C*.

In the mixel decomposition we focused on the two terms of:

$$R(i,v) = K_g G + \frac{C}{A}\iint_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds$$

The first term described how the sunlit background proportion proceeded to a maximum point as viewing and illumination positions in the hemisphere coincided. The second term described how the sunlit *S. damnosum s.l.*riverine larval habitat surface, composed of the Lambert Ian facets became maximally exposed to view at the hotspot, while those facets on tops became dominant at large viewing zenith angles. The hot spot correlation effect refers to the observed brightening which can occur when viewing a scene from the same direction as the solar illumination (Burrough and McDonnell, 1998) which for predictive, vector, insect habitat, predictive, risk modeling is commonly noted in the visible and NIR spectral regions (Jacob et al., 2011a).

We then analyzed how the first term K_gG varied with illumination and viewing geometry. As in Strahler and Jupp (1990), we assumed that the spatial object of interest (that is, *S. damnosum s.l.* habitat) and its associated explanatory sub-mixel spectral, predictor, covariate, coefficient estimates had the shape of a spheroid, with vertical half-axis equal to *b*, horizontal radius equal to *R* and a height to the center of the spheroid *h*. To accommodate the spheroidal shape in the derivations of the shadowed habitat areas, we used the transformation:

$$\theta' = \tan^{-1}\left(\frac{b}{R}\tan\theta\right)$$

We solved this equation by replacing θ with the angle that would generate the same shadow area for a sphere. For simplicity, we assumed that the centers of the spheroids were randomly distributed in depth from h_1 to h_2 over A. We then assumed that *G* and *C* were constant average signatures over Ag and Ac for properly endmember modelling Kg and Kc = A_c/A.

Next, the equation:

$$R(i,v) = K_g G + \frac{C}{A}\iint_{A_c} \frac{\langle i,s\rangle}{\cos\theta_i}\frac{\langle v,s\rangle}{\cos\theta_v} ds$$

was employed where K_g was expressed in a Boolean model and $K_g = e^{-\lambda\pi R^2\left[\sec\theta_i' + \sec\theta_v' - \overline{O}(\theta_i,\theta_v,\phi)\right]}$ where $\overline{O}(\theta_i,\theta_v,\phi,)$ was the average of the overlap function $O(\theta_i,\theta_v,\phi,h)$ between illumination and viewing shadows of the georeferenced capture point and its associated Precambrian rock and ripple water features. The Boolean model for a random subset of the plane or higher dimensions, analogously is a common tractable models in stochastic geometry (Cressie, 1993). Jacob et al. (2011c) used a Poisson point process of rate λ in the plane of a spectrally, decomposed, georeferenced, aquatic, larval, habitat,of *An. arabiensis,* and then made each sampled point be the center of a random set. The resulting union of the overlapping sets was a realization of the Boolean model [$\mathcal{B}$]. More precisely, the spectral parameter estimators were λ. Then a probability distribution on compact sets was created for each sampled point ξ employing a Poisson point process which used a set C_ξ from the distribution, and then defined $\mathcal{B}$ as the union $\cup_\xi(\xi + C_\xi)$ of the translated sets. To illustrate tractability with one simple formula, the mean density of $\mathcal{B}$ was then defined by a QuickBird sub-mixel endmember classification which revealed that $1-\exp(-\lambda \mathrm{E}\,\Gamma)$ where Γ denoted the sampled *S. damnosum s.l.* riverine habitat surface area of C_ξ.

In this research, ϕ was the difference in Azimuth angle between viewing and illumination positions of the QuickBird imaged objects associated to the *S. damnosum s.l.*, riverine, larval, habitat capture point. To simplify the equation, we approximated the overlap function by the overlap area and center positions of the ellipses. This approximation is justified when solar zenith and viewing zenith angles are not too large (Strahler and Jupp, 1990). In the case of long ellipsoidal shadows, however, this approximation could have overestimated the width of the capture point hotspot in the Azimuthal direction and underestimated the width of the hotspot in

the Azimuthal direction. To improve the accuracy and preserve the proper hotspot width information, we developed another approximation as follows. We used the equations $\phi = 0$ *or* $\phi = \pi$. First, we considered the overlap function in the principal plane. We used W $\phi = 0$ and π as the elliptical illumination estimates to determine if all the viewing shadows were aligned in the same direction. The overlap area was then approximated by an ellipse with one axis equal to the overlap length and the other equal to the capture point width encompassing the Precambrian rock and ripple water mixel spectrally decomposed radiance components which yielded:

$$O(\theta_i, \theta_v, \phi) = \frac{1}{2}\left[\sec\theta_i' + \sec\theta_v' - \frac{h}{b}\left|\tan\theta_i' - \tan\theta_v' \cos\phi\right| \right]$$

In the geometric-optical model, the shape of the hotspot function was based on the viewing and illumination positions in the model, and these diverged due to the shape and height of the spheroids. The "hotspot" was defined as a site with a combination of Precambrian rock and turbid water, sub-mixel, endmember data associated with *S. damnosum* s.*l.* larval breeding habitat. The equation *S* = (AgG + ACC + AZZ + AtT) IA, where S represented the reflectance of the QuickBird mixel; G, C, Z, and T were the reflectances of sunlit ground and shadowed ground and canopy, respectively; Ag through A were the corresponding areas of the four components; and A was the mixel size, and was helpful to understand how the physical shape of the *S. damnosum s.l.* riverine habitat, Precambrian rock and riffle water components governed the shape of the overlap functions. The exact overlap function on the principal cone was also obtained using the hybrid of the geometric optical model for capturing bi-directional reflectance values over the riverine larval habitat attributes.

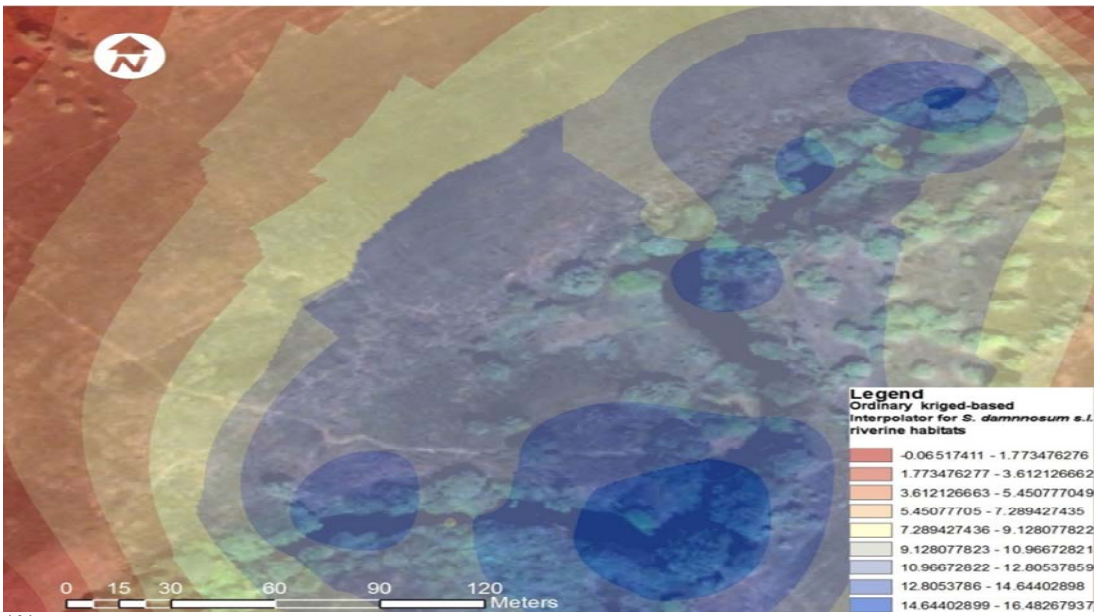
In the model, the viewing zenith was the viewing direction, which had a different Azimuth than the illumination position. Rather than computing the overlap of ellipses rendered from the *S. damnosum s.l.* habitat, the Precambrian rock and its riffle water spectral components at arbitrary inclinations and distances directly, we fit a linear function to the diminution of the overlaps generated from the model residuals using Azimuth angles. We approximated the Azimuthal cut off of the hotspot and linearly interpolated the model residuals. The residual output from the equation was then used to determine that; (1) the Azimuthal width of hotspot effect was basically determined by a ratio; (2) the outward width of hotspot on the principal plane was determined by ratio, and; (3) the inward width was determined by both. The composition of this signature was 34% red, 11% blue and 55% green. This signature corresponded to habitats consisting of fast flowing water over a base of Precambrian rock.

We then employed an ordinary kriged-based algorithm in ArcGIS Geostatistical Analyst for predicting other unknown, unsampled *S. damnosum s.l.* habitats at the study site using the reference signature generated from the canopy endmember extraction algorithms. For determining optimal explanatory predictor covariate coefficients, a variogram was constructed which expressed the variation in the spectral estimatators. In this research the variogram [that is, $2\gamma(x, y)$] was a function describing the degree of dependence between the predictedgeoreferenced S*. damnosum s.l.* riverinelarval habitats [i.e., $Z(x)$]. This was defined as the expected squared increment of the forecasted values between the forecasted georeferenced habitat locations. Our model was nonnegative since it was the expectation of a square. The covariance function was related to variogram by $2\gamma(x,y) = C(x,x) + C(y,y) - 2C(x,y)$. In this research, the $\gamma(x,y) = E(|Z(x) - Z(y)|^2)$ was equivalent to $\gamma(y,x)$ which was a symmetric function, consequently, $\gamma_s(h) = \gamma_s(-h)$ was an even function. In this research, the function was also a semivariogram as it was a conditionally negative definite function, (i.e., for all weights $w_1, \ldots, w_N$ subject to $\sum_{i=1}^{N} w_i = 0$) and the *S. damnosums.l.* habitat geolocations ($x_1, \ldots, x_N$); thus,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} w_i \gamma(x_i, x_j) w_j \leq 0$$

.

Semivariogram plot of the logit scale model residuals confirmed a short range spatial pattern up to a distance of approximately 20 km from the predicted *S. damnosum s.l.,*riverine, larval, habitat site. To carry out this process, residuals for all observed points were calculated on the logit ($\ln(p/1 - p)$) scale of the model. A kriged map of deviance residuals was then calculated which was added to the predicted values on the logit scale. Spatial dependence, displayed by these plots was analyzed using the semivariogram. The addition of kriged residuals allowed the maps to deviate from the model and move closer to the original sampled, canopy-related, predictor, covariate, coefficient, indicator, measurement values. These smoothed values improved the final maps of the forecasted *S. damnosum s.l.* habitats and its associated canopy cover, Precambrian rock and rippled water components sampled. An exponential model was then fitted to the semivariogram, using a range of 72.6 m, a nugget of 0.21 (variance), a lag size of 11.5 m with 12 lags and a partial sill of 0.24 (variance) (Figure 8). Thereafter, a predictive *S. damnosum* s.l. habitat canopy-based map was generated for a neighbouring riverine study site (Figure 9). The forecasted *S. damnosum s.l.* habitats were then field-verified which revealed a 100% correlation with the predicted estimates. We then used



(A)

(B)

Figure 9. (A) Kriged*S. damnosum s.l.* habitat pixel spectral reflectance estimates with predicted using a QuickBird endmember reference signature, (B) Predicted *S. damnosum s. l.* habitats in a neighbouring riverine ecosystem at the *S. damnosum s. l.* riverine epidemiological study site.

the ARIMA procedure outlined by Box and Jenkins for quantifying the nonlinear least squares iterations. Given our time-series dependent data where t was an integer index and X_t was the sampled immature *S. damnosum s.l.* data, then an ARMA (*p*,*q*) model was constructed based on $\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)X_t-\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$ (9) where L was the lag operator, the α_i were the parameters of the autoregressive part of the model, the θ_i was the parameters of the moving average part and the ε_t were

error terms. The error terms [that is, ε_t] were generally assumed to be i.d.d variables sampled from a normal distribution with zero mean. Assuming now that the polynomial $\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)$ has a unitary root of multiplicity *d*., we rewrote equation 9 as $\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)=\left(1-\sum_{i=1}^{p-d}\phi_i L^i\right)(1-L)^d$. The ARIMA (*p*,*d*,*q*) process expressed this polynomial factorization property, which was then rendered by: $\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$. Our model output resembled a particular case of an ARMA (*p*+*d*,*q*) process having the auto-regressive polynomial with some roots in the unity.

In terms of estimation methods, METHOD = ML option in PROC ARIMA which produced MLEs. The MLEs were computed by letting the univariate ARMA models be $\phi(B)(W_t-\mu_t)=\theta(B)a_t$ where a_t was an independent sequence of normally distributed innovations with mean 0 and variance σ^2. Here μ_t was the mean parameter μ plus the transfer function inputs. The log-likelihood function was then written as follows: $-\frac{1}{2\sigma^2}x'\Omega^{-1}x-\frac{1}{2}\ln(|\Omega|)-\frac{n}{2}\ln(\sigma^2)$. In this equation, *n* was the number of georeferenced *S. damnosum s. l.* spatiotemporal-sampled, riverine, spectral, larval, habitat observations, $\sigma^2\Omega$ was the variance of x as a function of the ϕ and θ parameters, and $|\Omega|$ denoted the determinant. The vector x was the time series W_t minus the structural part of the model μ_t, written as a column vector, as follows:

$$x=\begin{bmatrix}W_1\\W_2\\\vdots\\W_n\end{bmatrix}-\begin{bmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{bmatrix}$$

The MLE of σ^2 in the endemic, transmission-oriented, Euclidean, distance-based, transmission-oriented models was then computed as:

$$s^2=\frac{1}{n}x'\Omega^{-1}x$$

Note that the default estimator of the variance was divided by $n-r$, where *r* was the number of estimators in the model, instead of by *n*. Specifying the NODF option caused a divisor of *n* to be used. The log-likelihood with respect to σ^2 was then optimized using additive constants which then rendered $-\frac{n}{2}\ln(x'\Omega^{-1}x)-\frac{1}{2}\ln(|\Omega|)$ in the model residuals. We let H be the lower triangular matrix with positive elements on the diagonal such that $HH'=\Omega$. We also let e be the vector $H^{-1}x$ in the model. The concentrated log-likelihood with respect to σ^2 was then written as $-\frac{n}{2}\ln(e'e)-\ln(|H|)$ and $-\frac{n}{2}\ln(|H|^{1/n}e'e|H|^{1/n})$. The MLEs were then produced by using the LMA to minimize the following sum of squares: $|H|^{1/n}e'e|H|^{1/n}$ in the models. The subsequent analysis of the residuals was done by using e as the vector of residuals.

Thereafter the METHOD=CLS option produced robust conditional least squares (CLS) estimates. The series x_t was represented in terms of the seasonal-sampled *S. damnosum s.l.* riverine larval habitat observations, as follows:

$$x_t=\theta_t+\sum_{i=1}^{ov}\pi_i x_{t-i}$$

The π weights were computed from the ratio of the ϕ and θ polynomials, as follows:

$$\frac{\phi(B)}{\theta(B)}=1-\sum_{i=1}^{ov}\pi_i B^i$$

The CLS method produced estimates minimizing:

$$\sum_{t=1}^{n}\hat{a}_t^2=\sum_{t=1}^{n}\left(x_t-\sum_{i=1}^{ov}\hat{\pi}_i x_{t-i}\right)^2$$

Whereby $\hat{\pi}_i$ was computed from the estimates of ϕ and θ at each iteration. For attaining robust METHOD=ULS and METHOD=ML, initial estimates were computed using the METHOD=CLS algorithm. The arthropod-related, infectious disease, transmission–oriented, risk-model, residual estimates were then obtained by applying least squares MLE to the noise series. Thus, for transfer function models, the MLE option did not generate the full multivariate ARMA MLEs employing only the univariate likelihood function which in this research was applied to the noise series.

Because PROC ARIMA in SAS/GIS employed all of the available data for the input series to generate the noise series, other start-up options for the transferred series was implemented by prefixing the seasonal-sampled, *S. damnsoum s.l.,* riverine, larval, habitat observations as non-interpolated covariate coefficients. For example, we fit a transfer function model to the sampled, onchocerciasis, transmission-oriented, variable Y with the single input X. Thereafter, we employed a start-up using 0 for the seasonal sampled values by prefixing the coefficients to the actual data using an observation with a missing value for Y and a value of 0 for X. PROC ARIMA was then computed which generated the information criteria, (AIC). The AIC can be used to compare competing

models fit to the same series (Cressie, 1993). The AIC rendered $AIC = 2k - 2\ln(L) = 2k - 2(C - \chi^2/2) = 2k - 2C + \chi^2$ for each Euclidean, distance-based ArcGIS, endemic, delineated, transmission-oriented zone. Thereafter, the model order depicted the pattern associated with an ARMA series.

In this research, the chi-square statistics employed in the test for lack of fit were computed using the Ljung-Box formula:

$$\chi^2_m = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{(n-k)} \text{ where } r_k = \frac{\sum_{t=1}^{n-k} a_t a_{t+k}}{\sum_{t=1}^{n} a_t^2} \text{ and } a_t \text{ were the residual series.}$$

The Ljung–Box test is a type of statistical test of whether any of a group of autocorrelations of a series are different from zero (Griffith, 2003). Therefore, in this research instead of just testing randomness at each distinct lag, we also determined the "overall" randomness based on a number of lags (that is, a portmanteau test). A portmanteau test is a type of statistical hypothesis test in which the null hypothesis is well specified, but the alternative hypothesis is more loosely specified whereby tests constructed in this context can have the property of being at least moderately powerful against a wide range of departures from the null hypothesis (Cressie, 1993).

The extended, sample, autocorrelation, function method tentatively identified the orders of non-stationary ARMA process based on iterated least squares estimates of the autoregressive parameter estimators at each ArcGIS classified transmission zone at the study site. Given a stationary or non-stationary time series $\{z_t : 1 \le t \le n\}$ with mean corrected form $\tilde{z}_t = z_t - \mu_t$ with a true autoregressive order of $p+d$, a true moving-average order of q was estimated using the unknown orders $p+d$ and q by analyzing the autocorrelation functions associated with filtered series of the form:

$$w_t^{(m,j)} = \Phi_{(m,j)}(B)\tilde{z}_t = \tilde{z}_t - \sum_{i=1}^{m}\phi_i^{(m,j)}\tilde{z}_{t-i}$$

In this series B represented the backshift operator, where $m = p_{min}, \ldots, p_{max}$ were the autoregressive test orders, where $j = q_{min}+1, \ldots, q_{max}+1$ represented the moving-average test orders, and where $\phi_i^{(m,j)}$ were the optimized autoregressive parameter estimates under the assumption that the series was an ARMA (m,j) process. For purely predictive, autoregressive models $(j=0)$, OLS is used to consistently estimate $\phi_i^{(m,0)}$ (Cressie, 1993).

In this research, consistent estimates from the ARMA models were obtained also by the iterated least squares recursion formula, which was initiated by the pure autoregressive estimates:

$$\phi_i^{(m,j)} = \phi_i^{(m+1,j-1)} - \phi_{m+1-i}^{(m,j-1)}\frac{\phi_{m+1}^{(m+1,j-1)}}{\phi_m^{(m,j-1)}}$$

The jth lag of the sample autocorrelation function of the filtered series $w_t^{(m,j)}$ then extended the sample autocorrelation function, which in this research was denoted as $r_{\tilde{z}(m)} = r_{\tilde{z}}(w^{(m,j)})$ for each onchocerciasis, endemic transmission-oriented, risk-based, transmission zone. The standard errors of $r_{\tilde{z}(m)}$ were then computed in the usual way by using Bartlett's approximation of the variance based on the sample autocorrelation function:

$$var(r_{\tilde{z}(m)}) \approx (1 + \sum_{i=1}^{j-1} r_i^2(w^{(m,j)}))$$

The Barletts approximation calculates standard error with an approximation that was appropriate when the series represents a moving average process of order *k-1* (Hosmer et al., 2000). With this method, standard errors grew with increasing lags in the endemic, spatially, autoregressive model residuals. If the true model is an ARMA $(p+d,q)$ process, the filtered series $w_t^{(m,j)}$ follows an MA (q) model for $j \ge q$ so that $r_{\tilde{z}(p+d)} \approx 0 \quad j > q$ and $r_{\tilde{z}(p+d)} \ne 0 \quad j = q$ (Tsay and Tiao, 1984). In this research, we showed that the extended sample autocorrelation from endemic, transmission-oriented, risk–based, epidemiological model satisfied $r_{\tilde{z}(m)} \approx 0 \quad j-q > m-p-d \le 0$ and $r_{\tilde{z}(m)} \ne c(m-p-d, j-q) \quad 0 \le j-q \le m-p-d$ when $c(m-p-d, j-q)$ was a non-zero constant or a continuous random variable bounded by –1 and 1.

An extended sample autocorrelation function (ESACF) table was then constructed by $r_{\tilde{z}(m)}$ for $m = p_{min}, \ldots, p_{max}$ and $j = q_{min}+1, \ldots, q_{max}+1$ to identify the ARMA orders. The orders were tentatively identified by finding a right (that is, maximal) triangular pattern with vertices located at $(p+d,q)$ and $(p+d,q_{max})$ in which all elements were insignificant based on asymptotic normality of the autocorrelation function. The vertex $(p+d,q)$ thereafter identified the order associated with an ARMA (1,2) series.

The smallest canonical (SCAN) correlation method was then tentatively used to identify the orders of a stationary or non-stationary ARMA process in the endemic transmission-oriented model. We used simulation to study the efficacy of the modification for applying test statistics to analyze daily logSCAN for robust ARIMA

model selection in SAS/GIS. LogSCAN data can be applied to either non-transformed or differenced series (Tsay and Tiao, 1985). We then used the time series $\{z_t : 1 \le t \le n\}$ with mean corrected form $\tilde{z}_t = z_t - \mu_z$ with a true autoregressive order of $p+d$ employing a true moving-average order of q. We used the SCAN method to analyze eigenvalues of the correlation matrix of the ARMA process. Thus, for autoregressive test order $m = p_{min}, \ldots, p_{max}$ and for moving-average test order we had $j = q_{min}, \ldots, q_{max}$. We then let $Y_{m,t} = (\tilde{z}_t, \tilde{z}_{t-1}, \ldots, \tilde{z}_{t-m})'$. Then we computed the $(m+1) \times (m+1)$ matrix

$$\beta(m, j+1) = \left(\sum_t Y_{m,t-j-1} Y'_{m,t-j-1}\right)^{-1} \left(\sum_t Y_{m,t-j-1} Y'_{m,t}\right) \left(\sum_t Y_{m,t} Y'_{m,t}\right)^{-1} \left(\sum_t Y_{m,t} Y'_{m,t-j-1}\right) = \beta'(m, j+1) \hat{\lambda}^2(m, j) \beta'(m, j+1) \beta(m, j+1)$$

Where t ranged from $j+m+2$ to n. We found the smallest eigenvalue was $\hat{\lambda}^2(m, j)$, of $\hat{\lambda}^2(m, j)$ and its corresponding normalized eigenvector, $\Phi_{m,j} = (1, -\phi_1^{(m,j)}, -\phi_2^{(m,j)}, \ldots, -\phi_m^{(m,j)})$ in the model. The squared canonical correlation estimate was $\hat{\lambda}^2(m, j)$. Using the $\Phi_{m,j}$ as AR(m) coefficients, we obtained the residuals for $t = j+m+1$ to n, by using the formula:

$$u_t^{(m,j)} = z_t - \phi_1^{(m,j)} z_{t-1} - \phi_2^{(m,j)} z_{t-2} - \ldots - \phi_m^{(m,j)} z_{t-m}$$

From the sample autocorrelations of the residuals, $r_k(u)$, we approximated the standard error of the squared canonical correlation estimates using $var(\hat{\lambda}^2(m, j))^{1/2} \approx d(m, j) / (n - m - j)$ which rendered $d(m, j) = (1 + 2\sum_{k=1}^{j-1} r_k(u^{(m,j)}))$. The test statistic we employed was an identification criterion which was $c(m, j) = -(n - m - j) \ln(1 - \hat{\lambda}^2(m, j) / d(m, j))$. This expression was asymptotically χ^2 if $m = p+d$ and $j > q$ or if $m \ge p+d$ and $j = q$ in the transmission-oriented, epidemiological, risk model. We noticed that in the residual for $m > p$ and $j < q$, there was more than one theoretical zero canonical correlation between $Y_{m,t}$ and $Y_{m,t-j-1}$. Since the $\hat{\lambda}^2(m, j)$ were the smallest canonical correlations for each (m, j), the percentiles of $c(m, j)$ were less than those of a χ^2; A SCAN table was then constructed using $c(m, j)$ to determine which of the $\hat{\lambda}^2(m, j)$ were significantly different from zero. The ARMA orders were then tentatively identified for the model by finding a pattern in which the $\hat{\lambda}^2(m, j)$ were insignificant for all test orders $m \ge p+d$ and $j \ge q$ which was then depicted in SAS/GIS I (Figure 9).

DISCUSSION

In this research we robustly quantitated the seasonal-sampled georeferenced endemic risk-based explanatory predictor covariate coefficients and their uncertainty indicators (for example, latent autocorrelated error coefficients) within an spatially dependent geographically weighted matrix in PROC ARIMA. By so doing, we attained fine-tuned unbiased versions of random-walk and random-trend model specifications. In our pre-whitening approach the autocovariate parameter error estimators were estimated and removed from the data and the model, and the GLM was re-fitted. We employed the pre-whitening method of Cochrane and Orcutt (1949) which in this research was performed in SAS/GIS assuming that the errors generated from the estimated predictive residual variance followed a first-order autoregressive process. Pre-whitening is a preconditioning technique for the correlation analysis method (Cressie, 1993).

In this research our pre-whitening involved applying a filter to the input signal $u(k)$ and the output signal $y(k)$ to obtain a pre-whitened input signal $u'(k)$ and a pre-whitened output signal $y'(k)$. After calculation of a GLM, the amount of serial correlation was successfully estimated using pairs of successive residual estimated values (for example, ET, $ET+_1$). In our model the filter was well designed such that $u'(k)$ represented the white noise. By so doing, we also were able to successfully perform a correlation analysis on $u'(k)$ and $y'(k)$ to estimate the impulse response. The impulse response we estimated for the onchocerciasis, endemic, transmission-oriented risk model with $u'(k)$ and $y'(k)$ was equivalent to the impulse response estimate when the following equation remained true:

$$y'(k) = \sum_{n=0}^{\infty} u'(k-n) h(n) + e(k)$$

Therefore, in our predictive, autoregressive, spatiotemporal, arthropod-related risk model, $u'(k)$ was white noise. We then selected "ARIMA" as the model type to evaluate the order of non-seasonal differencing, and to set all the AR and MA terms to zero. We noticed that the seasonal change-related, parameter estimators in our model had stationary noise, suggesting that the mean (that is, constant) forecasting estimator had to be applied to accurately quantitate seasonal differences in the onchocerciasis–related, spatiotemporal-sampled,

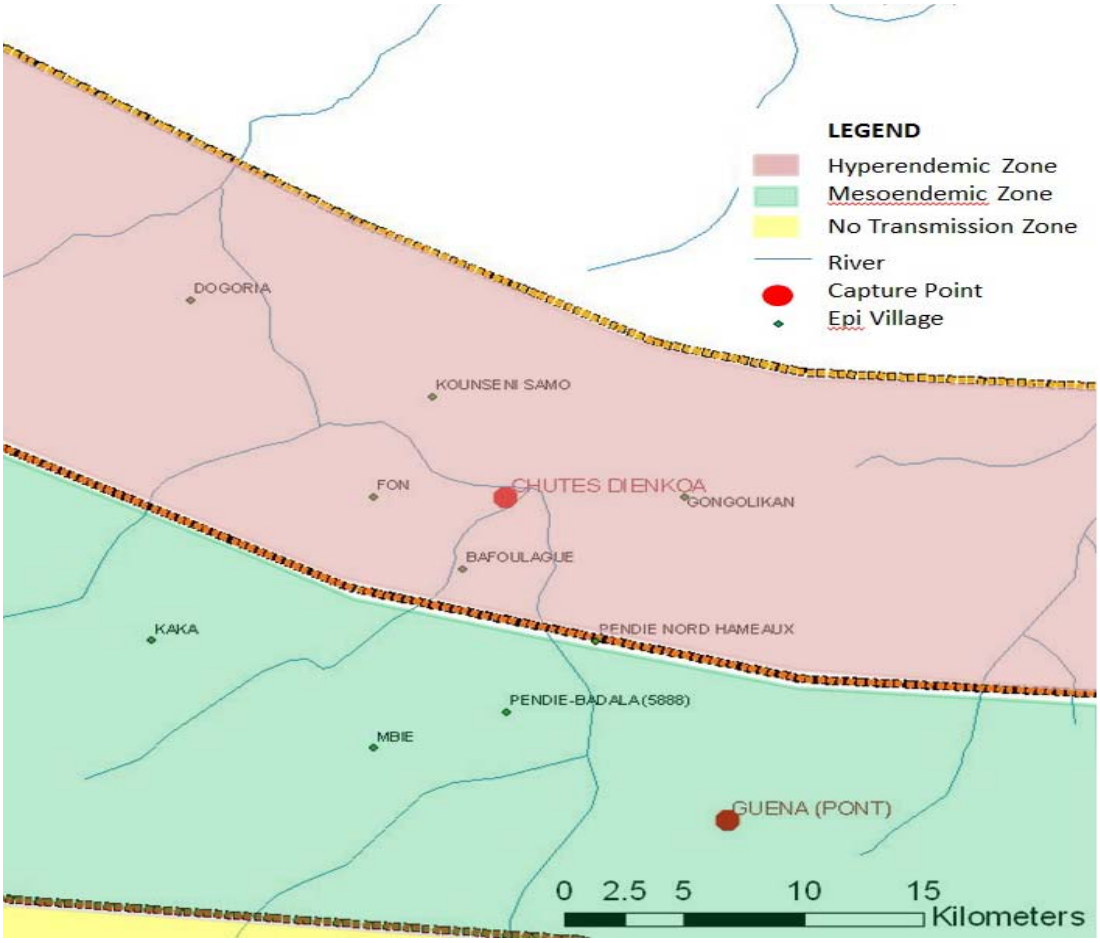


Figure 10. ARIMA predicted *S. damnosum s.l.*-related data feature attributes in riverine epidemiological study site.

epidemiological data. The ARIMA specifications then quantified geographically varying lags in the endemic risk-based distribution model employing the Euclidean-based distance measurements. The output revealed that the spatial data attribute features were positively spatially auto-correlated with each 5 km stratified geolocation.

Interestingly, the "mean" and the "constant" in risk-based distribution ARIMA model-fitting results were different numbers whenever the model included AR terms. Thus, in the future when fitting the ARIMA model to Y in which *p* is the number of autoregressive terms in an endemic, transmission-oriented, risk-based, onchocerciasis-related, distribution model, y can be used to denote the difference (that is, stationarized) version of Y- [for example, y(t) = Y(t)-Y(t-1)]. By so doing, only one non-seasonal difference would be required. Thereafter, the AR (p) forecasting equation for *y* in the model would be:

$$\hat{y}(t) = \mu + \phi_1 y(t-1) + \phi_2 y(t-2) + \ldots + \phi_p y(t-p)$$

Statistically speaking this equation would then be an ordinary multiple regression model where "mu" is the constant term, "phi-1" is the coefficient of the first lag of y, and so on. Thereafter, an infectious disease researcher or local program manager may be able to convert this slope-intercept form employing a predictive regression equation whose equivalent form represents in terms of deviations from the mean. Thus, by letting m denote the mean of the stationarized series y, a *p*-order autoregressive equation may be written in terms of deviations from the mean in their endemic transmission-oriented risk-based endmember, distribution model using the equation:

$$\hat{y}(t) - m = \phi_1 (y(t-1) - m) + \phi_2 (y(t-2) - m) + \ldots + \phi_p (y(t-p) - m)$$

Collecting all the constant terms in this equation would then allow identifying values rendered that are equivalent to the "mu" within a robust, onchocerciasis, endemic, transmission-oriented, predictive, regression-based, risk-based equation if:

$$\mu = m(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$$

In this research we actually estimated "m" along with the other endemic, transmission-oriented, risk-based, distribution model *p* estimators and reported this as the mean in the model-fitting results, along with its standard error and *t*-statistic. Thereafter, a constant (that is, "mu") was calculated according to the preceding formula [that is, constant = mean *(1 - sum of AR coefficients]. If an ARIMA model does not contain any AR terms, the MEAN and the CONSTANT are identical (Cressie, 1993).

Thereafter, in our model the estimated predictive residual variance had one order of non-seasonal differencing only which was the mean trend factor (for example, average period-to-period change). Our final predictive, robust, spatiotemporal, autoregressive, arthropod-related, risk-based, epidemiological, distribution model had one order of seasonal differencing only and the mean was the annual trend factor (for example, average year-to-year change).

To accurately spatially forecast the appropriate dataset of spatiotemporal, onchocerciasis, endemic, transmission-oriented, risk-based ARIMA distribution model residuals, thereafter we employed lengthy, time-series, dependent, *S. damnosum s.l.,* georeferenced, feature attributes to identify and predict, endemic, transmission zones at the riverine study site. We identified the order(s) of differencing needed to stationarize the sampled data series. We then removed the gross features of seasonality from the sampled data. Differencing is an excellent way of transforming a non-stationary series to a stationary one (Cressie, 1993). In this research this differencing was facilitated in conjunction with a variance-stabilizing transformation which we constructed by deflating the seasonal-sampled, time series-dependent, observational, explanatory, predictorsin ArcGIS geospatial analyst.

In applied statistics, a variance-stabilizing transformation is a data transformation that is specifically chosen either to simplify considerations in graphical exploratory data analysis or to allow the application of simple regression-based or analysis of variance techniques (Everitt, 2002). The aim behind the choice of our variance-stabilizing transformation was to find a simple function f to apply to the sampled, endemic, onchocerciasis, transmission-oriented, covariate coefficient (that is, x) in the empirical, ecological, spatiotemporal-sampled, dataset to create new sampled values $y = f(x)$ such that the variability of the sampled values y was not related to their mean value. We knew the explanatory, covariate, coefficient values x in our risk-based model data realizations from the different calculated Poisson distributions we generated had different mean values (that is, μ). Further, we noticed that the variance in our model varied with the mean. Commonly, the Poisson distribution has a variance that is identical to the mean (Haight, 1967). Fortunately a simple variance-stabilizing transformation where $y = \sqrt{x}$ linearly rectified the sampling variance associated with the spatiotemporal-sampled, observational, predictor, explanatory estimates. The estimates were nearly constant. Our methods were similar to that of the Anscombe transform which proposed a form of the square root transform $z = 2\sqrt{x + 3/8}$ aimed at stabilizing the variance of the Poisson distribution to a value of approximately 1 with the transformed distribution being approximately normal especially for larger model mean values (for example, m>20) (Anscombe, 1948).

In our riverine-based, predictive, autoregressive, epidemiological, risk models we assumed a mean value of zero for forecasting the hyperendemic risk regions at the riverine epidemiological study site. Our autoregressive, ARIMA, transmission-oriented, risk-based, time-series matrix contained monthly sampled feature attributes (for example, field-sampled larval habitat observational predictors), whose seasonal period was 12. The first difference of the seasonal difference at period *t* was then (Y(t) - Y(t-12)) - (Y(t-1) - Y(t-13). Applying the zero-mean forecasting model to this series yielded the equation: (Y(t) - Y(t-12)) - (Y(t-1) - Y(t-13)) = 0. Rearranging terms to put Y(t) by itself on the left, the equation then was Y(t) = Y(t-1) + Y(t-12) - Y(t-13) (10). For example, our spatiotemporal, onchocerciasis, endemic, transmission-oriented, risk-based data set we used September, 1987 to predict a seasonal, hyperendemic, *S. damnosum s.l.,* riverine–related, geolocational, transmission-oriented, explanatory predictor value of Y in October, 1987 by simply computing Y(Oct'1987) = Y(Sep'1987) + (Y(Oct'1986) - Y(Sep'1986) (11). In other words, October's hyperendemic transmission forecasts were equal to September's value plus the September-to-October regressed explanatory, observed, predictor, covariate, coefficient, indicator values from the previous sampled year in our model. Equivalently, equation 11 was rewritten as: Y(Oct'1987 = Y(Oct'1986) + (Y(Sep'1987) - Y(Sep'1986), therefore, the October's forecasts was equal to last October's sampled value plus the year-to-year change observed from the previous month. The preceding two equations then were mathematically identical; in actuality they were just rearranged terms on the right-hand-side.

Thereafter, a correlagram plotted the autocorrelation values in SAS/GIS for the time series at different lags (that is, "autocorrelation function"). The seasonal trend differences observed in a particular sampling period was the same or just a random step away from the trend that was observed from the previous sampling period in our model. Our ArcGIS-related forecasting risk model however revealed erroneous seasonal random trend model specifications. Even the most spatially accurate random walk or randomized, trend, seasonal, endemic,

Euclidean, distance-based, epidemiological model may contain undetected error residuals (for example, correlated disturbances) (Cressie, 1993). As such, we constructed a simple exponential smoothing model with a weighted moving average of past values to determine a predictive, regression-based, time series equation which was described as:

$$\hat{Y}(t) = Y(t-1) - \theta e(t-1)$$

where e (t-1) denoted the error at period t-1. The coefficient of the lagged forecast error was then written with a negative sign for reasons of mathematical symmetry while "Theta" in the predictive equation corresponded to the quantity "1-minus-alpha" in the exponential smoothing formulas. Note, our model resembled the predictive regressive-based equation for the ARIMA (1,1,0) model, except that instead of a multiple of the lagged difference the model included a multiple based on the lagged forecast error. Some non-stationary, time-series, dependent, non-seasonal, data feature attributes exhibit noisy fluctuations around a slowly-varying mean (Cressie, 1993). In other words, taking the most recent onchocerciasis–related, observation as the forecast of the next observation was better than using an average of the sampled observational predictors in order to filter out the noise for estimating a localized mean. Then theLjung–Box analyses tested whether any of a group of autocorrelations rendered from our time-series dependent data were different from zero. In this research the "portmanteau" test of Ljung-Box assessed the null hypothesis that a series of residuals exhibited no autocorrelation for a fixed number of lags *L*, against the alternative that some autocorrelation coefficient ρ (*k*), *k* = 1, ..., *L* was nonzero.

The test statistic was:

$$Q = T(T+2)\sum_{k=1}^{L}\left(\frac{\rho(k)^2}{(T-k)}\right),$$

where *T* was the sample size, *L* was the number of autocorrelation lags, and ρ (*k*) was the sample autocorrelation at lag *k*. Under the null, the asymptotic distribution of *Q* was the chi-square with *L* degrees of freedom. Instead of testing randomness at each distinct lag, our model effectively tested the overall randomness within a QuickBird-classified, Euclidean, distant-based, zone at the riverine study site based on a number of lags (that is, a portmanteau test). Values that equaled to 1 indicated rejection of the null of no autocorrelation in favor of the alternative. The qualitative observations from the eigen decomposition algorithm were then confirmed by the seasonal, random-walk, trend, model, dataset of residuals which revealed that seasonal hyeprendemic, stratified, epidemiological, prevalence rates and the ArcGIS derived Euclidean distance-based, algorithmic measurements were based on the 0 to 5 km transmission zone.

Thereafter, by letting y denote the differenced (that is, stationarized) version of Y in the autoregressive regression-based equation y(t) = Y(t)-Y(t-1), non-seasonal, differential variables were employed for targeting aggregations of prolific habitat locations (that is, hypernendemic area) based on spatiotemporal, field-sampled, count data. The AR(p) forecasting regression based, equation for y was $\hat{y}(t) = \mu + \phi_1 y(t-1) + \phi_2 y(t-2) + \ldots + \phi_p y(t-p)$. This was just an ordinary multiple, regression, risk model in which "mu" was the constant term, "phi-1" which was the coefficient of the first lag of y, and so on. Now, internally, the software (SAS/GIS) converted this slope-intercept form of the spatiotemporal, autoregressive, onchocerciasis, endemic, transmission-oriented, risk–based, predictive, regression equation to an equivalent form in terms of the model's deviations from the mean. Thereafter, by letting m denote the mean of the stationarized series y in the endemic, transmission-oriented, regression-based, model residuals, the *p*-order autoregressive, predictive, time series equation was re-written in terms of the deviations from the mean as:

$$\hat{y}(t) - m = \phi_1 (y(t-1) - m) + \phi_2 (y(t-2) - m) + \ldots + \phi_p (y(t-p) - m)$$

By collecting all the constant terms in this equation, the residuals were then determined to be equivalent to the "mu" form of the equation:

$$[\ \mu = m(1 - \phi_1 - \phi_2 - \ldots - \phi_p)\ ]$$

Conclusion

ArcGIS determined all distance-based measurements from the georeferenced *S. damnosum s.l*. riverine capture point using an Euclidean allocation algorithm. The algorithm rendered the direction from each cell to the closest georeferenced stratified village source. Poisson regression models prioritized the covariates at each demarcated ArcGIS-derived transmission zone. A spectral unmixing analyses then identified the spectral S. *damnosum s.l.,* habitat, canopy, endmember, sub-mixel components that encompassed illumination geometrical values reflected from the Precambrian rock and ripple water at the vertices of the simplex. The final endmember model output identified the correct fractional presence of each canopy-oriented, endmember, spectral, component, emitted from the georeferenced *S. damnosum s.l.,* habitat capture point and its associated Precambrian rock and rippled water. Thereafter, a kriged smoothed map displayed the spatial patterns of all productive *S. damnosum s.l.* habitats based on the endmember reference signature. By so doing, we were able to define the distance at which 95% of the sill was reached for the

asymptotic variogram model.

Multiple terms were caused by coefficient quantization, interpolation and linear-approximation error, respectively. The estimates where then analyzed with an eigenvector spatial filtering algorithm using a positive-definite covariance matrix in SAS/GIS® which removed the spatial covariate dependence by partitioning the original sampled endemic, transmission-oriented data feature attributes and into two synthetic variates: (1) a filter variate capturing latent spatial dependency and (2) a non-spatial variate that was free of spatial dependence. The space-time autocorrelation estimation actually involved two specifications: one casting a percentage at a sampled, prevalence, stratified, village, geographical location as a function of the preceding *in situ*, entomological-related, prevalence value as well as the preceding neighboring sampled villages values, a lagged specification; and the other casting a percentage at location as a function of the preceding *in situ* value as well as the contemporaneous neighboring values. The geographic distribution of the sampled explanatory co-variate coefficients based on the immature *S. damnosum s. l* counts exhibited positive spatial autocorrelation in all models tested: like larval counts aggregated in geographic space. After modeling the misspecification term explicitly, the remaining residuals become white noise. This allowed us to calibrate the autoregressive, endemic, transmission-oriented, landscape, risk-based models with the standard OLS estimation procedure which then revealed that the hyperendemic region was from 0 (georeferenced capture point) to 5 km from the capture point. This research demonstrated that the eigenvector spatial filtering approach can be embedded into a semi-parametric statistical framework for spatially targeting endemic onchocerciasis transmission zones.

ACKNOWLEDGEMENT

This work was produced by the US National Institute of Health/Fogarty International Center under Unnasch T.U. SR01TW008508.

Conflict of interest

The author declared he has no conflict of interest.

REFERENCES

Abramowitz M, Stegunl A (1972). *H*andbook of mathematical functions with formulas, graphs, *a*nd mathematical tables. 10th printing, with corrections. ed. United States National Bureau of Standards Applied mathematics series, Washington: U.S. Govt. Print. Off. xiv, 1046 p.

Burrough PA, McDonnell R (1998). Principles of geographical information systems. Spatial information systemsOxford ; New York: Oxford University Press.

Cochrane D, Orcutt GH (1949). Application of least squares regression to relationships containing auto-correlated error terms. J. Am. Stat. Assoc. 44:32-61.

Cressie NAC (1993). Statistics for spatial data. Wiley Series in probability and mathematical statistics Applied probability and statistics. New York: Wiley.

Goodchild MF (1980). *St*atistics in Geography - a Practical Approach - Ebdon, D. Geograp. l Anal.12:411-412.

Griffith DA (1996). *S*patial autocorrelation and eigenfunctions of the geographic weights matrix accompanying georeferenced data. Can. Geogr. 4:351-367.

Griffith DA (2000). Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. Linear Algebra Appl. 321:95-112.

Griffith DA (2003). Spatial autocorrelation on spatial filtering Springer.

Griffith DA (2005). A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. Int. J. Health Geogr. 4:18-26..

Haight FA (1967). Handbook of the Poisson Distribution New York: John Wiley & Sons.

Jensen JR (2005). Introductory digital image processing a remote sensing perspective, 3rd ed. Upper Saddle River, N.J.: Prentice Hall.

Ross SM (2007). Introduction to probability models. 9th ed, Amsterdam, Boston: Academic Press.

Schowengerdt RA (2007). Remote sensing : models and methods for image processing, 3rd ed. Oxford: Academic Press.

Sondow J, Zudilin W (2006). *E*uler's constant, q-logarithms, and formulas of Ramanujan and Gosper. Ramanujan J. 12:225-244.

Strahler AH, Jupp DLB *(*1990). Modeling bidirectional reflectance of forests and woodlands using Boolean models and geometric optics. Remot. Sens. Environ. 34:153-166.

Tsay RS, Tiao GC (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. J. Am. Stat. Assoc. 79:84-96.

Tsay RS, Tiao GC (1985). Use of Canonical-Analysis in Time-Series Model Identification. Biometrika 72:299-315.

Whittle P (1954). On Stationary Processes in the Plane. Biometrika 41:434-449.