Quasi-likelihood techniques in a logistic regression equation for identifying *Simulium damnosum* s.l. larval habitats intra-cluster covariates in Togo

Benjamin G. JACOB*, Robert J. NOVAK, Laurent TOE, Moussa S. SANFO, Abena N. AFRIYIE, Mohammed A. IBRAHIM, Daniel A. GRIFFITH and Thomas R. UNNASCH

*Global Infectious Disease Research Program, Department of Public Health, College of Public Health, University of South Florida, 3720 Spectrum Blvd, Suite 304, Tampa, FL 33612, USA; Onchocerciasis Control Programme (OCP), B. P. 1473 Avenue Norte, Zombo, Sector 8, Ouagadougou, Burkina Faso, West Africa; Department of Chemistry, University of Pittsburgh, 3260 Forbes Avenue, First Floor, USA; Department of Biology, Wake Forest University, 226 Winston Hall, Box 7325 Reynolda Station, Winston-Salem, NC 27109, USA; School of Economic and Policy Sciences, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021, USA

(Received 15 September 2011; final version received 27 June 2012)

The standard methods for regression analyses of clustered riverine habitat data of *Simulium damnosum* s.l. a major black-fly vector of onchocerciasis, postulate models relating observational ecological-sampled parameter estimators to prolific habitats without accounting for residual intra-cluster error correlation effects. Generally, this correlation comes from two sources: (1) the design of the random effects and their assumed covariance from the multiple levels within the regression model and (2) the correlation structure of the residuals. Unfortunately, inconspicuous errors in residual intra-cluster correlation estimates can overstate precision in forecasted *S. damnosum* s.l. riverine larval habitat explanatory attributes regardless how they are treated (e.g. independent, autoregressive, Toeplitz, etc.). In this research, the geographical locations for multiple riverine-based *S. damnosum* s.l. larval ecosystem habitats sampled from two preestablished epidemiological sites in Togo were identified and recorded from July 2009 to June 2010. Initially, the data were aggregated into PROC GENMOD. An agglomerative hierarchical residual cluster-based analysis was then performed. The sampled clustered study site data was then analyzed for statistical correlations using monthly biting rates (MBR). Euclidean distance measurements and terrain-related geomorphological statistics were then generated in ArcGIS. A digital overlay was then performed also in ArcGIS using the georeferenced ground coordinates of high and low density clusters stratified by annual biting rates (ABR). The data was overlain onto multitemporal sub-meter resolution satellite data (i.e. QuickBird 0.61 m wavbands). Orthogonal spatial filter eigenvectors were then generated in SAS/Geographic Information Systems (GIS). Univariate and nonlinear regression-based models (i.e. logistic, Poisson, and negative binomial) were also employed to determine probability distributions and to identify statistically significant parameter estimators from the sampled data. Thereafter, Durbin–Watson statistics were used to test the null hypothesis that the regression residuals were not autocorrelated against the alternative that the residuals followed an autoregressive process in AUTO-OREG. Bayesian uncertainty matrices were also constructed employing normal priors for each of the sampled estimators in PROC MCMC. The residuals revealed both spatially structured and unstructured error effects in the high and low ABR-stratified clusters. The analyses also revealed that the estimators, levels of turbidity, and presence of rocks were statistically significant for the high-ABR-stratified clusters, while the estimators distance between habitats and floating vegetation were important for the low-ABR-stratified cluster. Varying and constant coefficient regression models, ABR-stratified GIS-generated clusters, sub-meter resolution satellite imagery, a robust residual intra-cluster diagnostic test, MBR-based histograms, eigen decomposition spatial filter algorithms, and Bayesian matrices can enable accurate autoregressive estimation of latent uncertainty affects and other residual error probabilities (i.e. heteroskedasticity) for testing correlations between georeferenced *S. damnosum* s.l. riverine larval habitat estimators. The asymptotic distribution of the resulting residual adjusted intra-cluster predictor error autocovariate coefficients can thereafter be established while estimates of the asymptotic variance can lead to the construction of approximate confidence intervals for accurately targeting productive *S. damnosum* s.l. habitats based on spatiotemporal field-sampled count data.

**Keywords:** *Simulium damnosum* s.l. cluster covariates; QuickBird; Onchocerciasis; Annual biting rates; Bayesian; Togo

1. Introduction

Onchoceriasis or river blindness is a human filarial infection which causes blindness in infected people and is transmitted by *Simulium* species or black flies. Traditionally, the designation of at-risk communities for onchocerciasis is accomplished through intensive ground-based epidemiological surveys of communities located in rural riverine areas in which the disease is endemic. In 1974, the World Health Organization began the Onchocerciasis Control Programme (OCP) in 11 West African countries. Thereafter, extensive
countrywide field-sampled surveys began in 1975 (1). These surveys revealed that onchoceriasis was hyperendemic with prevalence rates around 70% with some areas showing human blindness rates up to 9%. The frequency and level of evolution of ocular lesions and of onchocercal blindness were among the highest in the world. The data were employed by the OCP to justify and deploy aerial applications of larval insecticides to reduce the populations of *Simulium damnosum s.l.*, the primary vector of onchoceriasis, thereby, reducing transmission to humans (2). *Simuliidae* or black flies in the *S. damnosum* Theobald complex are the only insect vectors of human onchoceriasis in West African countries (3).

The original OCP vector control efforts resulted in a dramatic reduction in the transmission of onchoceriasis, an accomplishment that was maintained for over 12 years (4). The effectiveness of the Program is illustrated by the reduction in annual biting rates (ABRs) of *Simulium* species at Pont frontiere on the Leraba River, which declined from 26 in 1975 to 3 in 1989, when the control program ended. Similarly, at Loabe on the Nakambi River and at Ziou Zabre (both in Burkina Faso), the ABR fell from 6090 and 11,879 in 1975 to 238 and 1465 in 1989, respectively. Significant downward trends as a result of the black fly control program were observed at Bagre on the Nazinon River and at Bitou on the Nauhao River, a tributary of the Nakambi River. In contrast, certain areas within the 11 countries covered by the OCP, such as the Oti River tributaries in Togo, Leraba/Comoe Rivers in Burkina Faso, Baoulé and Sankaran Rivers in Mali, White Bandama in Côte d’Ivoire and the Black Volta River region in Ghana, the effectiveness of vector control measures was not as successful.

Within the 11-country OCP area the movement of black flies was found to represent a significant impediment to the success of vector control efforts, particularly in areas on the eastern and western border. Dry season reinvasion of black flies traveling on Harmattan winds or West African trade winds from the north was blamed for the rapid dispersion of insecticide-resistant black flies to other river basins (5). As a result of this repopulation, the boundaries of the program area were extended to include additional riverine breeding sites that were serving as the reservoir for these migrating flies (6).

Macrogeographic factors such as the complexity of the landscape, human population movements, and the suspension of larviciding of neighboring black fly producing rivers may also have had an effect on eliminating parasite reservoirs. Additionally, marked differences between spatial-temporal-sampled habitats with respect to microgeographical distribution such as larval habitat production, host preferences, vectorial capacity, and susceptibility to larvicides, biting cycle, and population age structure may also have complicated larvicidal operations. It has been suggested that spatial-temporal patterns of vector insect larval habitat production are driven by two mechanisms, namely, (i) variation in intrinsic properties of breeding habitats, which affect growth and survival of immature populations and (ii) the spatial locations of focal habitats in relation to human habitation (2). These habitat variabilities require vigorous quantitative statistical analyses for implementing control programs. By doing so, varying linear outputs and discrete-time state-space models could be used to remotely target productive *S. damnosum s.l.* larval habitats based on spatiotemporal field-sampled count data. Treatments or habitat perturbations should be based on surveillance of larvae in the most productive areas of an ecosystem (7).

In this paper we constructed multiple spatiotemporal cluster-based autoregressive residual error matrices employing Durbin–Watson (DW) first-order autocorrelation statistics, spatial filter orthogonal eigenvectors and Bayesian hierarchical generalized linear mixed models to spatially target productive *S. damnosum s.l.* larval habitats based on field-sampled count data in two riverine epidemiological breeding study sites in Togo. Since contagious processes, such as conspecific attraction and others can generate time series-dependent error patterns in *S. damnosum s.l.* riverine larval habitats species abundance that cannot be explained by simple residual hierarchical cluster-based regression models, we assumed by combining linear and nonlinear residual predictive-error estimation algorithms we could qualitatively assess and quantify varying and constant intra-cluster regression-based disturbances (e.g. conditional heteroskedasticity and serial error correlation).

The importance of this research may also be expressed in the GIScience literature regarding representations of geographic space as well as various literatures concerned with time series-dependent vector insect larval habitat modeling. Presently, there is not a steady current of literature on representation issues in GIS for remotely quantitatively assessing spatiotemporal-sampled vector insect data. Thus, the potential synergy of developments in GIS, spatial statistics, and entomology may not be apparent to researchers in their respected disciplines. The fusing of these independent research trajectories into a common cohesive agenda (e.g. GIS/remote integrated vector aquatic larval habitat control-based cyber environment) could generate geostatistical tools that might reveal new insights into the roles of physical and human geography in vector borne disease transmission. Results from the entomological and epidemiological surveillance activities have indicated fly infectivity levels and infection in humans that require improved programme (1). Therefore, the objectives of this research were to: (1) perform a hierarchical regression-based cluster analyses using multiple georeferenced *S. damnosum s.l.* larval habitat parameter estimators, (2) construct multiple stepwise linear models using the sampled explanatory variables, (3) filter all latent serial autocorrelation error coefficients using an eigenfunction decomposition algorithm, (4) formulate a customized uncertainty diagnostic test using the random effects from an iterative Bayesian analyses, and (5) validate all forecasted estimates using a cumulative residual analyses for qualitatively assessing and
2. Material and methodology

2.1. Study site

Five onchoceriasis/black fly sites located in Togo were used in this study including: Mo, Landa Pozada, Bagan, Titra, and Sarakawa-Kpleou. The riverine epidemiological study sites are located approximately 100 km of Kara, a city in northern Togo, situated in the Kara Region, 413 km north of the capital Lomé. The Haugeau River flows a little way south of Kara and is the main resource of water for the region. North of Kara is the Oti River which runs through a sandstone plateau. This area is vegetative savanna and is characterized by granite and gneiss outcrops. The Oti River drains the plateau and is a main tributary of the Volta River. The rivers and its tributaries are characterized by a period of flooding from July to November with a peak in September and a lengthy low water period from January to June. Land uses include arable land, permanent crops, permanent pastures, forests, and woodland. The climate is generally tropical with average temperatures about 30 °C (86 °F) in the region during the black fly season lasting up to seven months, while the dry desert winds of the Harmattan blow south from November to March, bringing cooler weather.

2.2. Remote sensing data

In this research, we used Landsat Enhanced Thematic Mapper Plus (ETM+) data for remotely determining geographic locations of the S. damnosum s.l. larval habitat clusters based on multiple regression-based parameter estimators. The Landsat ETM+ image data consisted of eight spectral bands with a spatial resolution of 30 m for bands 1–5 and band 7. Resolution for band 6 (thermal infrared) was 60 m and resolution for band 8 (panchromatic) was 15 m. The total scene size we used was 170 × 183 km. QuickBird (www.digitalglobe.com) satellite images were also acquired for the riverine epidemiological breeding study site areas. We acquired 11-bit data in five spectral bands covering panchromatic (525–924 nm), blue (447–512 nm), green (499–594 nm), red (620–688 nm), and near-infrared (NIR) (755–874 nm) wavelengths for the study sites. At nadir, the nominal ground sample distance was 0.61 m (panchromatic) with a nominal swath width of 16.5 km. The basic products were delivered at the native sensor resolution and swath width of the image acquisition. The products were then resampled to a panchromatic ground resolution of 0.61 m and cropped to define geographic polygons of the riverine epidemiological study sites.

The satellite imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS Imagine V.8.7™. The QuickBird scene size per sampled S. damnosum s.l. riverine study site image was 25 km².

2.3. ABR cluster-based classification

Initially, a 5 km buffer was placed around the riverine epidemiological breeding sites using the Landsat Thematic data in ArcGIS. The field and remote explanatory predictor covariate coefficient estimates encompassing the georeferenced breeding sites were also entered into SAS 9.2® (Carey, North Carolina). FLEXIBLE/FLE in SAS was then used to request the flexible-beta method. The PROC CLUSTER statement started the procedure which specified a residual robust hierarchical clustering algorithm employing METHOD = FLEXIBLE. Stratified random sampling was then performed using PROC FREQ and PROC SURVEYSELECT. A routine was developed to select the stratified samples by ABR rates (Figure 1). The final model revealed that the highest density ABR-based cluster was the Bagan riverine breeding study site while the lowest ABR cluster was the Sarakawa-Kpleou study site.

2.4. Habitat mapping

Field sampling was then conducted in the Bagan and Sarakawa-Kpleou riverine epidemiological breeding study sites from July 2009 to July 2010. The study sites were mapped and classified using a CSI-Wireless Differentially Corrected Global Positioning Systems (DGPS) Max receiver. This remote technology employed an OmniStar L-Band satellite signal yielding a positional error of .179 m (±.392 m). We then overlaid the georeferenced spatiotemporal-sampled S. damnosum s.l. larval habitat regression-based parameter estimators onto the QuickBird data in ArcGIS. We placed a robust digitized grid-based algorithm onto the satellite data to generate efficient spatial sampling units. Once overlaid, the ArcGIS grid-based data files consisted of columns and rows of uniform cells coded according to the parameter estimator ground coordinates sampled in each epidemiological study site. Each grid cell within the matrix contained an environmental-sampled attribute value as well as sampled black-fly habitat location-based geocoordinates (Figure 2).

The Bagan and Sarakawa-Kpleou epidemiological riverine study sites were then examined extensively using longitude, latitude, and altitude data. This criterion involved attaining the centrographic measures of spatial mean, distance between the sampled georeferenced larval habitats and the distance from sampled site to the nearest human habitation, for qualitatively assessing the sampled items within the selected clusters in ArcGIS. In this research the sampled habitat data were comprised of individual georeferenced observations together with a battery of categorical explanatory variables which were expanded into multiple attribute measures using histograms (Figure 3).
2.5. Environmental parameters
Distance measures were also recorded in ArcGIS spatial analyst as Euclidean distances. The nearest source was determined by the Euclidean Distance function in ArcGIS®. The Euclidean direction output raster contained the azimuth direction. The Euclidean Allocation function identified the nearest human habitation center closest to each digitized grid cell. This weighted function assigned space between the sampled larval habitats. These geometric distances in multidimensional space were computed as: distance \((x, y) = \left( \sum_i (x_i - y_i)^2 \right)^{1/2} \). All observations sampled in this research are listed in Table 1.

2.6. Spatial ecohydrological model
Three-dimensional models of the riverine epidemiological breeding study sites were then constructed based on digital elevation model (DEM) statistics (Figure 4). The latest version of PCI Geomatics Orthoengine® software was used to construct the models from the \textit{S. damnosum s.l.} estimators. We generated the DEMs from stereo data which required the use of geometric models and the DGPS ground coordinates of the sampled riverine larval habitats.

2.7. Regression analyses
Logistic regression models were then constructed using a 95% confidence level to ascertain whether the proportions of the sampled estimators in each riverine epidemiological study site differed by sampled larval habitat geolocations. In this research, the SAS procedures PROCMIXED were used to fit the linear larval habitat models. The regression models assumed independent
Bernoulli outcomes denoted by $Y_i = 0$ or 1, taken at the sampled larval habitat sites (e.g. $i = 1, 2, \ldots, n$). In probability and statistics, a Bernoulli process is a finite or infinite sequence of binary random variables, so theoretically it is a discrete-time stochastic process that takes only two values, canonically 0 and 1 (8). The indicator values were then described by $X_i$, a 1-by-$(K+1)$ vector of $K$ values, and a 1, for the intercept term, which represented a sampled larval habitat site geolocation in each study site. The probability of a 1 being realized for
Figure 3. Histogram of MBR and distance to the nearest human habitation in the Togo study site.

Table 1. Environmental-sampled cluster-based S. damnosum s.l parameter estimators sampled in the Bagan and Sarakawa-Kpleou riverine breeding sites as entered in SAS®.

| Variable | Description                      | Units               |
|----------|----------------------------------|---------------------|
| GCP      | Ground control points            | Decimal-degrees     |
| FLOW     | Flowing water                    | Presence or absence |
| HIGHT    | Height of water                  | Formazin turbidity unit |
| TURB     | Turbidity of water               | Percentage          |
| AQVEG    | Aquatic vegetation               | Percentage          |
| HGVEG    | Hanging vegetation               | Percentage          |
| DDVEG    | Dead vegetation                  | Percentage          |
| RCKS     | Rocks                            | Percentage          |
| MMB      | Man-made barriers                | Type (e.g. dams, bridges) |
| DISHAB   | Distance between habitats        | Meters              |

Figure 4. A DEM based on georeferenced S.damnosum s.l. riverine habitats for the Sarakawa-Kpleou breeding study site.
the binary outcome data was then given by
\[ P(Y = 1 | X_i) = \exp(X_i\beta) / (1 + \exp(X_i\beta)), \]
where \( \beta \) was the \((K + 1)\)-by-1 vector of nonredundant parameter estimators and \( P(Y = 0 | X_i) = 1 - P(Y = 1 | X_i) \).

A Poisson regression with statistical significance was also calculated by a 95% confidence level in SAS. We used the PROC GLIMMIX of SAS to fit the sampled \( S. \) damnosum s.l. estimators in each study site. In our Poisson model, it was assumed that the dependent variable \( Y \) had a Poisson distribution given the independent variables \([X_1, X_2, \ldots, X_m, P(Y=k) = e^{-\mu}(\mu^k/k!)} \), where the log of the mean \( \mu \) was assumed to be a linear function of the independent variables. That is, \( \log(\mu) = \text{intercept} + b_1 X_1 + b_2 X_2 + \cdots + b_3 X_3 \) implied that \( \mu \) was the exponential function of the independent variables, where \( \mu = \exp(\text{intercept} + b_1 X_1 + b_2 X_2 + \cdots + b_3 X_3) \). The regression model then was rewritten in the following form: \( \log(\mu(n)) = \log(N) + \text{intercept} + b_1 X_1 + b_2 X_2 + \cdots + b_3 X_3 \), where \( n \) was the total number of georeferenced habitat samples harvested in each study site. The log of variable \( n \) was then used as an offset. By doing so, a regression-dependent parameter estimator with a constant coefficient of 1 was incorporated into the models. The log of the incidence, \( \log(\mu/n) \), was then quantified as a linear function of the independent variables in each model. The maximum likelihood method was then used to estimate the observational coefficients derived from the Poisson residuals.

In this research, the parameter estimator \( \lambda_i(X_i) \) was both the mean and the variance of the Poisson distribution based on the regressed predictor covariate coefficients sampled at each riverine study site. The analyses assumed independent counts (i.e., \( n_i \)), taken at the sampled habitat locations \( i = 1, 2, \ldots, n \), in each study site. That is, our Poisson regression assumed the response variable \( Y \) had a Poisson distribution and also assumed the logarithm of its expected value was modeled by a linear combination of the sampled parameter estimators in each study site. This expression was then written more compactly as \( \log(E(Y|x)) \) where \( x \) was an \( n + 1 \)-dimensional vector consisting of \( n \) independent variables concatenated to 1 and \( \theta \) which in actuality was simply a linearly linked to \( b \). Thus, in our Poisson residuals, \( \theta \) was an input vector \( x \) and the predicted mean of the associated Poisson distribution rendered from the sampled riverine lake habitat data in which our models was expressed as \( E(Y|x) = e^{tx} \).

Thereafter, the sampled estimators were denoted by matrix \( X_0 \), a \( 1 \times p \) which was based on the vector of the coefficient measurement indicator values for a specific sampled riverine habitat location \( I \) in each study site. A variance-stabilizing transformation was also performed to allow the analysis of application of variance techniques to the models. The aim behind our variance-stabilizing transformation was to find a simple function \( f \) to apply to the sampled estimator values (i.e., \( x \)) in the ecological data-sets to create new larval habitat values employing \( y = f(x) \) such that the variability of the sampled values (i.e., \( y \)) was not related to their mean value. While variance-stabilizing transformations are well known for certain parametric families of distributions, such as the Poisson and the binomial distribution, some types of data analysis proceed more empirically by searching among power transformations to find a suitable fixed transformation (9). Alternatively, if a time series-dependent data analysis suggests a functional form for the relation between variance and mean, this can be used to deduce a variance-stabilizing transformation (10).

Further, in the regression models the expected value of these data was provided by \( \mu(X_i) = n(X_i) \exp(X_i\beta) \) where \( \beta \) was the vector of the nonredundant Poisson estimators which was provided by \( \lambda_i(X_i) = \mu(X_i)/n_i(X_i) \). Thereafter, the models took the form \( \log(E(Y|x)) = a_x + b \). By doing so, a regression-dependent residual estimates in each study site. The logarithm of variable \( k \) was then used as an offset. By doing so, a regression-dependent parameter estimator with a constant coefficient of 1 was incorporated into the models. The log of the incidence, \( \log(\mu/n) \), was then quantified as a linear function of the independent variables in each model. The maximum likelihood method was then used to estimate the observational coefficients derived from the Poisson residuals.

In this research, the Poisson models were generalized by introducing an unobserved heterogeneity term for the sampled observations \( i \). Thus, the data was assumed to differ randomly in a manner that was not fully accounted for by the estimates rendered. These distributions were then reformulated as \( E(y_i|x_i, \tau_i) = \mu_i \), where the unobserved heterogeneity term \( \tau_i = e^{\tau_i} \) was independent of the vector of regressors \( X_i \); thus, the distribution of \( y_i \) was conditional on \( X_i \) and \( \tau_i \) was Poisson with a conditional variance of \( \mu_i \tau_i \). Then, the distribution \( f(y_i|x_i) \) was no longer conditional on \( \tau_i \) in \( x \) in both models. By doing so, linear-dependent residual estimates were then obtained by integrating \( f(y_i|x_i) \) with respect to \( \tau_i \): \( f(y_i|x_i) = \int_0^\infty f(y_i|x_i, \tau_i)g(\tau_i)d\tau_i \).

The regression residuals also revealed that the \( S. \) damnosum s.l. riverine larval habitat data attributes contained a constant term. As such, it was necessary to assume that \( E(e^{\mu_i}) = E(\tau_i) = 1 \) in order to identify the mean of the distributions. We had assumed that \( \tau_i \) followed a gamma (\( \theta, \theta \)) distribution in the models with \( E(\tau_i) = 1 \) and \( \tau_i = 1/\theta \). Then, the distribution \( f(y_i|x_i) \) was no longer conditional on \( \tau_i \) in \( x \) in both models. By doing so, linear-dependent residual estimates were then obtained by integrating \( f(y_i|x_i) \) with respect to \( \tau_i \): \( f(y_i|x_i) = \int_0^\infty f(y_i|x_i, \tau_i)g(\tau_i)d\tau_i \).

Unfortunately, extra-Poisson variation was detected in the variance estimates of the larval habitat models. Evidence of overdispersion indicates inadequate fit of the
Poisson model (9). A common way to deal with overdispersion for count data is to use a generalized linear model framework, where the most common approach is a “quasi-likelihood,” matrix with Poisson-like assumptions (i.e. quasi-Poisson) or a negative binomial model (11). As such, we constructed robust negative binomial regression models in PROC GLIMMIX with nonhomogeneous gamma distributed means by incorporating $x = \frac{b}{a} (x > 0)$ in equation (2.1) as in Jacob et al. (10).

The distribution was then rewritten as $f(y|x_i) = \frac{E(y^+)}{\gamma} \left( \frac{1}{\gamma + 1} \right)^{\gamma} \left( \frac{\gamma + 1}{\gamma} \right)^{\gamma + 1} e^{\gamma - 1} \left( \frac{y}{\gamma} \right)^{\gamma + 1}$, where $y = 0, 1, 2, \ldots$. The negative binomial distribution was then derived as a gamma mixture of the Poisson random variables. In both models, the conditional mean was $E(y|x_i) = \mu_i = e^{\beta \tilde{y}}$ and the conditional variance was $v(y|x_i) = \mu_i [1 + \sum_{j=1}^n \tilde{y}] = \mu_i [1 + z \tilde{y}]$. To further quantify the regression residuals, we specified DIST=NEGBIN($p=1$) in the MODEL statement in PROC REG. The negative binomial model NEGBIN1, set $p = 1$, then revealed the variance function $v(y|x_i) = \mu_i + z \mu_i^2$ was linear in the mean in both model residuals. The log-likelihood function of the NEGBIN1 models was then provided by $\varphi = \sum_{i=1}^N \left\{ \frac{\gamma - 1}{\gamma + 1} \ln(1 + z \tilde{y}) \right\}$, where $\tilde{y} = 0$. The gradient for the model by computing with $\frac{\partial \varphi}{\partial \tilde{y}} = \sum_{i=1}^N \left\{ \frac{\gamma - 1}{\gamma + 1} \ln(1 + z \tilde{y}) \right\}$ and $\frac{\partial \varphi}{\partial \mu} = \sum_{i=1}^N \left\{ -\frac{\gamma - 1}{\gamma + 1} \ln(1 + \tilde{y}) \right\}$. 

In this research, the negative binomial regression models with variance function $v(y|x_i) = \mu_i + z \mu_i^2$, were referred to as the NEGBIN2 model. To estimate these models, we specified DIST=NEGBIN ($p=2$) in the MODELS statement. A test of the Poisson distribution was then performed by examining the hypothesis that $z = \frac{1}{2} = 0$. A Wald test of this hypothesis was also provided which we used to report t statistics for the regression residuals. The log-likelihood functions of the models (NEGBIN2) was then rendered by the equation $\varphi = \sum_{i=1}^N \left\{ \frac{\gamma - 1}{\gamma + 1} \ln(1 + z \tilde{y}) \right\}$, where $\tilde{y} = 0$. The gradient for the model by computing with $\frac{\partial \varphi}{\partial \tilde{y}} = \sum_{i=1}^N \left\{ \frac{\gamma - 1}{\gamma + 1} \ln(1 + \tilde{y}) \right\}$ and $\frac{\partial \varphi}{\partial \mu} = \sum_{i=1}^N \left\{ -\frac{\gamma - 1}{\gamma + 1} \ln(1 + \tilde{y}) \right\}$.

2.8. DW statistics

Thereafter, we constructed a first-order autoregressive AR(1) error framework. The AR(1) models were constructed using the georeferenced estimators sampled from the Bagan and Sarkawa-Kpleou riverine breeding study sites. Each model was defined as $X_t = c + \sum_{j=1}^n \varphi_i X_{t-j} + \varepsilon_t$, where $\varphi_1, \ldots, \varphi_n$ was a sampled observation at a particular study site, $c$, was a constant, and $\varepsilon_t$ was white noise.

In this research, the AR(1) process was provided by $X_t = c + \varphi X_{t-1} + \varepsilon_t$ in the models where $\varepsilon_t$ was a white noise process with zero mean and variance $\sigma^2$. Thereafter, we used classifications, where, the autoregressive parameter processes were defined as wide-sense stationary, if $|\varphi| < 1$. This value was obtained as the output of stable filters whose input was white noise. The predictive autoregressive riverine larval habitat models was then denoted by $\mu$, thus, $E(X_t) = E(c) + \varphi E(X_{t-1}) + E(\varepsilon_t)$ such that $c$ and $\varphi$ and $c$. Additionally, in the residuals if $c$ was equal to $0$, then the mean was $0$. The variance was then delineated $var(X_t) = E\left\{ \left( X_t - \mu \right)^2 \right\} = \sigma^2$, where $\sigma^2$ was the standard deviation of $\varepsilon_t$. This was revealed by noting that $var(X_t) = \sigma^2 var(X_{t-1}) + \sigma^2$. Further, we noted that for $B_0 = E\left\{ \left( X_t - \mu \right)^2 \right\} = \sigma^2$, the autocovariance function in the residuals decayed with a constant time as defined by $\tau = -1/\ln(\varphi)$ in the models. Thus, in order to further define the autocorrelation function, we wrote $B_0 = K\varphi^0$ where $K$ was independent of $n$. We noted that $\varphi^0 = e^{-\lambda \ln(\varphi)}$. We then matched this value to the exponential decay law $e^{-\lambda \ln(\varphi)}$. We noticed that the spectral density function rendered was the Fourier transform of the autocovariance function in both models. The Fourier transform is a mathematical operation that decomposes a signal into its constituent frequencies (9). In this research, the discrete-time Fourier transform in the models was

$$\Phi(w) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} B_n e^{-\omega n} = \frac{1}{\sqrt{2\pi}} \left( 1 - \varphi^2 - 2\varphi \cos(w) \right)$$

This expression was periodic due to the discrete nature of the $X_t$ which was manifested as the cosine term in the denominators. We assumed that the sampling time (i.e. $\Delta t = 1$) was much smaller than the decay time (i.e. $\tau$) in the models. By doing so, we were then able to use a continuum approximation values to $B_n, B(t) \approx \frac{\varphi}{1 - \varphi^2 \cos^2(w)}$ which yielded a Lorentzian profile for the spectral density equation which was then delineated using $\Phi(w) = \frac{1}{\sqrt{2\pi} \varphi^0 \cos(w)}$ where $\gamma = 1/\tau$ was the angular frequency associated with $\tau$. In vector insect larval habitat modeling, Lorentz distribution is closely related to the Poisson kernel, which is the fundamental solution for the Laplace equation in the upper half-plane (10). In this research the Lorentz distribution had the probability density function

$$f(x; \gamma) = \frac{1}{\pi \gamma} \left[ \frac{(x-x_0)^2}{(x-x_0)^2 + \gamma^2} \right]$$
where \( x_0 \) was the sampled \( S. \) \textit{damnosum s.l.} predictor covariate coefficients specifying the geolocation of the peak of the distribution where \( \gamma \) was the scale estimator which specified the half-width at half-maximum. In our Lorentz distribution \( \gamma \) was also equal to half the interquartile range (i.e. the probable error in the model residuals).

Further, an alternative expression for \( X_t \) was derived from the epidemiological riverine larval habitat models by first substituting \( c = \varphi X_{t-1} - \mu_{t-1} \) for \( X_{t-1} \) in the defining equations. Continuing this process \( n \) times yielded \( X_t = c \sum_{k=0}^{N-1} q^k X_{t-N} + \sum_{k=0}^{N-1} q^k \varepsilon_{t-k} \). We noticed that for \( n \) approaching infinity, \( \varphi^N \) our model residuals approached zero and \( X_t = \frac{\varepsilon_{t-N}}{\varphi^N} + \sum_{k=0}^{\infty} q^k \varepsilon_{t-k} \).

The residuals also revealed that \( X_t \) was white noise convolved with the \( q^k \) kernel plus the constant mean. If the white noise \( \varepsilon_t \) is a Gaussian process then \( X_t \) is also a Gaussian process in a robust predictive vector insect larval habitat cluster-based regression model \((10)\). The model residual estimates also revealed that \( X_t \) was normally distributed when \( \varphi \) was close to one \( i \).

The DW statistic was then spatially derived for each model to detect the relationships between the sampled riverine larval habitats values separated from each other by a given time lag based on the forecasted uncertainty estimates from the regression analysis. We used the DWPROB option in SAS to print the significance level (i.e. \( p \)-values) for the DW tests. The DW statistic then tested the null hypothesis \( H_0 : \varphi_t = 0 \) against \( H_1 \sim \varphi_t \neq 0 \) in both models. In this research, the generalized DW statistic was written as:

\[
\text{DW}_{\text{g}} = \frac{\sum_{t=0}^{T} \hat{\varepsilon}_t \hat{A}_t^\prime \hat{\varphi}}{\sum_{t=0}^{T} \hat{\varepsilon}_t^2}
\]

where \( \hat{\varepsilon}_t \) was a vector of ordinary least square (OLS) residuals and \( \hat{A}_t \) was a \((T - j) \times T\) matrix. The generalized DW statistic \( \text{DW}_{\text{g}} \) was then rewritten as:

\[
\text{DW}_{\text{g}} = \frac{\text{YW}_{\text{g}} \text{AMT}}{\text{YW}_{\text{g}}} = \frac{\text{YW}^i \text{AMT}}{\text{YW}^i}
\]

where \( \text{YW}^i \) was \( L \)-times the model residuals and \( \text{AMT} \) was the upper triangular matrix \((10)\).

The marginal probability for the DW statistic in the models was then estimated as:

\[
\text{Pr}(\text{DW}_{\text{g}} \cap c) = \text{Pr}(\hat{\varphi} > 0) \quad \text{where} \quad \hat{\varphi} = \frac{\text{YW}^i \text{AMT}}{\text{YW}^i}.
\]

The marginal probability for the generalized DW statistic was computed by numerical inversion of the characteristic function \( \phi(\varphi) \) using the trapezoidal rule approximation to the marginal probability \( \text{Pr}(\hat{\varphi} > 0) \) was then specified by:

\[
\text{Pr}(\hat{\varphi} > 0) = \frac{1}{2} \left[ 1 - \sum_{k=0}^{K} \left| \text{Im}(\phi(\varphi + i\Delta)) \right| \right] + \text{E}_1(K) + \text{E}_1(\Delta)
\]

where \( \text{Im}(\phi) \) was part of the characteristic function and \( \text{E}_1(\Delta) \) and \( \text{E}_1(K) \) were integration factors and truncation errors, respectively. The trapezoidal rule is a way to calculate the definite integral \((9)\).

A numerically efficient algorithm was then used to quantify the error components in the first-order autoregressive models. To do so required \( O(N) \) operations for the evaluation of the characteristic function \( \phi(\varphi) \) in each model. In this research, the characteristic function in each autoregressive \( S. \) \textit{damnosum s.l.} larval habitat model was denoted by:

\[
\phi(\varphi) = [1 - 2i\varphi Q_1 - C_n^{-1}]^{-1} |x^2 - |x'x^2|^{-1/2} |x^2| |
\]

where \( \nu = (1 + 2i\varphi) \) and \( 1 - 2i\varphi Q_1 A_j \) and \( i = \sqrt{-1} \). Thereafter, by applying the Cholesky decomposition to the complex matrix \( V \), we obtained the lower triangular matrix \( G \) which satisfied \( V = GG^T \) in both models. Cholesky decomposition is a decomposition of a symmetric, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose \((11)\). The characteristic function then evaluated \( O(N) \) operations in the models by using:

\[
\phi(\varphi) = |G|^{-1} |x^2 - |x'x^2|^{-1/2} |x^2| |
\]

In AUTOREG, two alternative statistics (i.e. Durbin \( h \) and \( t \) were also employed to test for time varying autoregressive residual uncertainty coefficients that may have been asymptotically equivalent in the models. The DW tests are not valid when the lagged dependent variable is used in the regression model such that, the Durbin \( h \)-test or Durbin \( t \)-test can be used to test for first-order autocorrelation \((11)\). For the Durbin \( h \)-test, we specified the name of the lagged dependent variable in the LAG-DEP option. The \( h \) statistic was then written as:

\[
h = \hat{\rho} \sqrt{\frac{\sum_{t=0}^{T} \hat{\varepsilon}_t^2}{\sum_{t=0}^{T} \hat{\varepsilon}_t^2}}
\]

where \( \hat{\rho} = \frac{\sum_{t=0}^{T} \hat{\varepsilon}_t \hat{A}_t}{\sum_{t=0}^{T} \hat{\varepsilon}_t^2} \) and \( \hat{\varepsilon} \) was the least squares variance estimate for the coefficient of the lagged dependent variables. Durbin’s \( t \)-test consists of regressing the OLS residuals \( \hat{\varepsilon} \) on explanatory observational variables and \( \hat{\varepsilon}_{t-1} \) for quantifying the significance of the estimate for coefficient of \( \hat{\varepsilon}_{t-1} \) \((11)\).

In PROC AUTOREG, an estimation method was then used to construct multiple first-order autoregressive error matrices using the Yule-Walker (YW) method. The equation defining the \( AR \) processes in the models was constructed using \( X_t = \sum_{i=0}^{\rho} \varphi_i X_{t-i} + \varepsilon_t \). The YW equations we used included \( \gamma_m = \sum_{i=0}^{m} \varphi_i \gamma_{m-i} + \sigma^2 \delta_{m,0} \), where \( m = 0, \cdots, p \), which yielded \( p + 1 \) equations; where \( \gamma_m \) was the autocorrelation function of \( X, \sigma_{\varepsilon} \) was the standard deviation of the input noise process; and \( \delta_{m,0} \) was the Kronecker delta function. The Kronecker’s delta, in a robust vector insect larval habitat regression distribution model is a function of two sampled independent observations, usually integers where 1 is represented as 1 and everything above 1 and 0 is equal to 0 \((10)\).

Because the last part of our equations was nonzero only when \( m = 0 \), the first-order error estimation
equations were solved by representing the sampled riverine larval habitat data as a matrix for \( m > 0 \), thus,
\[
\begin{bmatrix}
\varphi_1 \\
\varphi_2 \\
\varphi_3 \\
\vdots
\end{bmatrix} = \begin{bmatrix}
\gamma_1
\gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_i \\
\gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_i \\
\gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_i \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma_i & \gamma_{i-1} & \gamma_{i-2} & \cdots & \gamma_0
\end{bmatrix}
\] solved all \( \varphi \).

Further, for \( m = 0 \) the model rendered
\[ \gamma_0 = \sum_{j=1}^{n} \varphi_j \gamma_{-j} + \sigma^2 \] which subsequently allowed us to solve \( \sigma^2 \). The full autocorrelation function was then derived by recursively calculating \( \rho(\tau) = \sum_{j=1}^{n} \varphi_j \rho(\tau - j) \) in the estimators. In this research, the YW equations were \( \gamma_1 = \varphi_1 \gamma_0 + \varphi_2 \gamma_{-1} \) and \( \gamma_2 = \varphi_1 \gamma_1 + \varphi_2 \gamma_{-1} \) when \( \gamma_{-k} = \gamma_k \). The equations then yielded \( \rho_1 = \gamma_1/\gamma_0 = \frac{\sigma^2}{\sigma^2 + \gamma_0} \) and the recursion formula rendered \( \rho_1 = \gamma_1/\gamma_0 = \frac{\sigma^2}{\sigma^2 + \gamma_0} \) in the residual variance.

The equations defining the AR processes in the models were then defined using \( X_l = \sum_{j=1}^{n} \varphi_j X_{l-j} + \epsilon_l \). Thereafter, we multiplied both sides by \( X_{l-m} \) and impute the expected values which yielded \( E[X_l X_{l-m}] = E[\sum_{j=1}^{n} \varphi_j X_{l-j} X_{l-m}] + E[\epsilon_l X_{l-m}] \) in the models. We noted that \( E[X_l X_{l-m}] = \gamma_m \) was the autocorrelation function in both models. The values of the model variance function in the models were independent on each other and \( X_{l-m} \) was independent of \( \epsilon_l \) when \( m \) was greater than zero. The autoregressive estimates for \( m > 0 \), \( E[\epsilon_l X_{l}] = 0 \) in the model residuals. Further, we noted that when \( m = 0 \), \( E[\epsilon_l X_{l}] = 0 \) for \( m \geq 0 \). This equation also rendered \( \gamma_m = E[\sum_{j=1}^{n} \varphi_j X_{l-j} X_{l-m}] + \sigma^2 \delta_m \) when \( m \geq 0 \). Thereafter, we employed \( \gamma_m = E[\sum_{j=1}^{n} \varphi_j X_{l-j} X_{l-m}] = \sum_{j=1}^{n} \varphi_j E[X_l X_{l-m}] + \sigma^2 \delta_m \) which can be computed using regression estimators and \( \Sigma \) which can then delineate \( \sigma^2 \) \( (I) \). Given \( \Sigma \) the efficient estimates of the \( S. damnosum \) s.l. larval habitat regression parameters, \( \beta \) was computed using generalized least square (GLS) for both models. The GLS yielded the unbiased estimate of the variance \( \sigma^2 \) in the model residuals.

The calculation of \( \gamma \) from the S. damnosum s.l. larval habitat AR analyses was complicated as it was completely dependent on the number of sampled observations in both models. Instead of actually calculating \( v \) and performing GLS in the usual way, we used a Kalman filter algorithm to transform the sampled data and compute the GLS results through a recursive process. The Kalman filter, also known as a linear quadratic estimation, is an algorithm that uses a series of measurements observed over time, containing noise (i.e., random variations) and other inaccuracies, which produces estimates of unknown variables that tend to be more precise than those that would be based on a single measurement alone \( (II) \). More formally, in this research, the Kalman filter operated recursively on streams of the noisy input riverine larval habitat data to produce statistically optimal estimators. The Shapiro-Wilk test was then used to test the null hypothesis that the sampled riverine larval habitat estimators \( x_1, \ldots, x_n \), came from a normally distributed population.

2.9. Bayesian analyses
We then used PROC MCMC for generating the multivariate density functions in a Bayesian estimation matrix. In PROC MCMC, we used the logarithmic of LOGMPDFWISHART for determining the Wishart distribution and, thereafter, the logarithm LOGMPDFWISHART for attaining a robust inverted-Wishart distribution. We let \( x \) be an \( n \)-dimensional random vector with mean vector \( \mu \) and covariance matrix \( \Sigma \). The density in the models was then \( pdf(x, \mu, \Sigma) = \frac{1}{\sqrt{2\pi}^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} tr(\Sigma^{-1} x) \right) \) with \( \mu' \in \mathbb{R}^n \). The trace of the square matrices \( A \) is then rewritten as: \( \text{tr}(A) = \sum_{i=1}^{n} a_{ii} \mu_i = 2\Sigma_{\mu}(\bar{\mu}) = \mu' \Sigma \mu \). The density function from the inverse-Wishart distribution was \( pdf(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|} \Gamma_n\left(\frac{n}{2}\right)} |x|^{-\frac{n+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma^{-1} x) \right) \) for \( \mu \neq 0 \), and \( D_n(\mu) = 2\sum_{i=1}^{n} \Gamma_i\left(\frac{n+1}{2}\right) \) in both models.

The marginal and conditional distributions from the inverse-Wishart distributed matrices were then further evaluated using \( A \sim W^{-1}(\mu, m) \). We partitioned the matrices for determining if \( \psi \) was compatible with each other using: \( A = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix} \) and \( \psi = \begin{bmatrix}
\psi_{11} & \psi_{12} \\
\psi_{21} & \psi_{22}
\end{bmatrix} \) where \( A_{ij} \) and \( \psi_{ij} \) were \( p \times p \) matrices. We then determined if: (i) \( A_{11} \) was independent of \( A_{12} \) and \( A_{22} \), \( A_{22} = A_{22} \) \( A_{21} \) \( A_{12} \) was the Schur complement of \( A_{11} \); (ii) \( A_{11} \sim W^{-1}(\psi_{11}, m - p) \); (iii) \( A_{11} \sim W^{-1}(\psi_{11}, m - p) \); (iv) \( A_{22} \sim W^{-1}(\psi_{22}, n - p) \). We used the conjugate distribution to make inference about a covariance matrix \( \Sigma \) for each model whose prior had a \( W^{-1}(\psi_{11}, m) \) distribution. We assumed if the sampled riverine larval habitat observations \( X = x_1, \ldots, x_n \) were independent \( p \)-variate Gaussian variables drawn from a georeferenced distribution pattern, then the conditional distribution had a \( W^{-1}(A + \psi, n + m) \) when \( A = XX^T \) was n times the sample covariance matrix.
Because in this research, the prior and posterior distributions were the conjugate to the multivariate Gaussian derived from the sampled estimators in both models.

In this research the probability density function was

\[ p(B \mid \Psi, m) = \frac{\| \Psi \|^{m/2} |B|^{m/2+p+1/2} \exp(-t r(\Psi^{-1}/2))}{2^{m/2} \Gamma[(m/2)]} \]

where \( \Psi = \Sigma^{-1} \) and \( \Gamma_p(\cdot) \) was the multivariate gamma function. The multivariate gamma function, [i.e. \( \Gamma_p(\cdot) \)], is a generalization of the gamma function which is useful in multivariate statistics which commonly appears in the probability density function of the inverse-Wishart distributions \( (\Pi) \). We also noticed that the gamma function had two equivalent expressions in both model estimates. One was \( \Gamma_p(a) = \int_{S > 0} \exp(-\text{trace}(S))|S|^{a-p-1/2} dS \), where \( S \) was positive-definite. The other one, \( \Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(a + (1 - p)/2) \), from which we determined the recursive relationships in the \( S_{\text{damnosum s.l.}} \) riverine larval habitat parameter estimators sampled at the epidemiological riverine study sites when \( \Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma(a + (1 - p)/2) \). Thus, in this research \( \Gamma_2(a) = \Gamma_2(a) = \pi^2 \Gamma(a) \Gamma(a - 1/2) \) and \( \Gamma_3(a) = \pi^3 \Gamma(a) \Gamma(a - 1/2) \).

2.10. Model validation

The residual riverine larval habitat parameter estimators were then validated using weighted cumulative models. The approach was implemented following the line of \( \text{geo}-\text{spatial Information Science} \) research. Initially, a test statistic using a \( \chi^2 \)-distribution was used. However, this statistic does not approximate the permutation test's \( p \)-values; thus, indicating that the within cluster-based regression residual estimates were conditional on our spatiotemporal-sampled data attributes. We assumed that the outcome could be verified using

\[ Y_i = X_i \beta + \epsilon_i, \quad \epsilon_i \sim (0, \sigma^2 / w_i) \]

where \( \beta \) was a \( p \times 1 \) vector of the regression-based parameters (i.e. \( \sigma^2 \)) which itself was an unknown variance parameter when \( w_i > 0 \). The weights assigned to in the validation models were \( i \) (\( i = 1, \ldots, n \)). The error terms, \( \epsilon_i \), was then independent with \( n_0 \) and variance \( \sigma^2 / w_i \) in the cluster-based regression residuals.

In the validation models, the weights, \( w_i \), represented the extra regional variability in \( Y_i \). We estimated \( \beta \), and \( \sigma^2 \) by \( \hat{\sigma}^2 \) employing \( \hat{U}_p(\beta) = \sum_{i=1}^{n} \hat{U}_i = \sum_{i=1}^{n} 1^n X_i T (Y_i - X_i \beta) / \hat{\sigma}^2 = 0 \) and \( \hat{U}_p(\sigma) = \sigma^2 - 1/n \sum_{i=1}^{n} 1^n w_i (Y_i - X_i \beta)^2 = 0 \) which simultaneously solved both equations. These conjectures were derived assuming \( \epsilon_i \sim (0, \sigma^2 / w_i) \). From these explanatory error estimating equations the residuals, \( \hat{\epsilon}_i = Y_i - X_i \hat{\beta} \) was used to test the ABR-stratified cluster-based larval habitat predictor covariate coefficient patterns in the Bagan and

Sarakawa-Kpleou riverine epidemiological study sites for quantifying higher than expected sum of residuals. Sum of residuals is a natural test statistic to use for regression validation as it has a defined distribution, and it has monotonic properties such that areas with higher sum of residuals can indicate areas with higher than expected outcomes \( (9) \). Thereafter, a two-dimensional moving block process was used over the forecasted locations of the prolific larval habitats employing \( (x_1, x_2) \), \( Z_{\text{loc}}(x_1, x_2) \) in each riverine study site which rendered

\[ Z_{\text{loc}}(x_1, x_2|b) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i(x_1, x_2|b) \hat{\epsilon}_i \]

where \( W_i(x_1, x_2|b) = I (x_1 - b < x_2 \leq x_1 + b, \ x_3 - b < r_i \leq x_3 + b) \) and \( w_i \) was a weighted indicator function.

3. Results

Our monthly biting rates (MBR)-related histograms revealed that September had the highest MBR rates \( (1800) \) at approximately 5.5 km while June, July, and October had MBR rates of 1100–1200 MBR measured at 3.5 km. The lowest MBR was in April at 100 MBR at a distance less than 1 km. Low to zero levels of transmission was measured inside villages less than 1 km distant from the riverine-based variations in the vectorial efficiency of the fly populations \( (2) \) (Figure 3).

In this research, ANOVA provided a test of whether or not the means of the high and low density ABR within cluster-based regression residual estimates were equal. Our \( F \)-test's \( p \)-values did not approximate the permutation test's \( p \)-values; thus, indicating that the within residual cluster-based observational data did not have the same effect in the high and low ABR-clusters. Power analyses were then performed. The test provided the probability of rejecting a false null hypothesis (i.e. a Type II error) in both models. As power increases, the chances of a Type II error decrease \( (11) \). In this research, the probability of a Type II error was the false negative rate \( (\beta) \); therefore, power in the larval habitat models was equal to \( 1 - \beta \). In Beta error probability sampling the power of a test is defined as \( 1 - \beta \) \( (11) \). The ANOVA contained repeated measures factors. The repeated measure design controlled for subject heterogeneity between the individual sampled larval habitat differences in the high and low ABR-stratified within residual cluster-based varying and constant predictor covariate coefficients.

We then tested for serial error correlation with lagged dependent variables in the models using the AUTOREG procedure for generating the generalized DW tests from the sampled estimators. Initially, we used the equation

\[ Y = X \beta + \nu, \quad X \text{ was an } n \times k \text{ data matrix, } \beta \text{ was a } k \times 1 \text{ covariate coefficient vector and } \nu \text{ was a } n \times 1 \text{ disturbance vector.} \]

The error term \( \nu \) was assumed to be derived by the \( j \)-th order autoregressive process:
\( v_l = \hat{e}_l - \phi_l v_{l-1} \) where \(|\phi_l| \leq I \). \( \hat{e}_l \) was a sequence of independent normal error terms generated from the explanatory estimators using a mean of 0 and the variance \( \sigma^2 \). We then used the DW statistic to test the null hypothesis \( H_0: \phi_1 = 0 \) against \( -H_1: -\phi_1 \neq 0 \). This process revealed that when the generalized DW statistic was:

\[
\bar{d}_l = \frac{\sum_{i=1}^{n} (v_i^2 - v_{i-1}^2)}{\sum_{i=1}^{n} \xi_{ij}^2} \text{ was OLS residuals in both models. We then used the matrix notation, }
\]

\[
d_l = \frac{YMAM}{YM} \quad \text{where } M = I_Y - X(X'X)^{-1}X' \text{ and } A_j \text{ constituted a } (N - j) \times N \text{ matrix: }
\]

\[
A_j = \begin{bmatrix}
-1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
0 & -1 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 \\
\end{bmatrix}
\]

which only existed in the models when \( j - 1 \) zeros were between -1 and 1 in each row of matrix \( A_j \). The test revealed that the QR factorization of the design matrix yielded a \( n \times n \) orthogonal matrix when \( Q \sim X = QR \) and when \( R \) was an \( n \times k \) upper triangular matrix. The tests also revealed there existed \( n \times (n-k) \) sub-matrices of \( Q \) such that \( Q_lQ_l' = M \) and \( Q_l'Q_l = I_{n-k} \) in the residual estimates. Consequently, the generalized DW statistic was stated as a ratio of two quadratic forms:

\[
d_l = \frac{\sum_{i=1}^{n} \hat{\psi}_i^2}{\sum_{i=1}^{n} \xi_i^2} \quad \text{where } \lambda_1 \cdots \lambda_n \text{ when the upper } n \text{ eigenvalues of } MA_j'M \text{ and } \xi_1 \text{ was a standard normal variate, and } n = \min(N - k, N - j) \text{. These eigenvalues were obtained by a singular value decomposition of } QA_j \text{. The singular value decomposition of an } m \times n \text{ real or complex matrix } M \text{ is a factorization of the form } M = U \Sigma V^\ast \text{ where } U \text{ is an } m \times m \text{ real or complex unitary matrix, } \Sigma \text{ is an } m \times n \text{ diagonal matrix with nonnegative real numbers on the diagonal, and } V^\ast \text{ (i.e. the conjugate transpose of } V \text{) is } n \times n \text{ real or complex unitary matrix} (H_i) \text{. In this research, the diagonal entries } \Sigma_{ii} \text{ of } \Sigma \text{ in the model estimates were the singular values of } M \text{. The } m \text{ columns of } U \text{ and the } n \text{ columns of } V \text{ were then the left singular vectors and right singular vectors of } M \text{, respectively in each riverine epidemiological larval habitat model.}

In this research, the marginal probability for \( d_l \), rendered \( c_0 \), in both model residual variance estimations procedures were quantified using

\[
\text{Prob} \left( \frac{\sum_{i=1}^{n} \hat{\psi}_i^2}{\sum_{i=1}^{n} \xi_i^2} \cap c_0 \right) = \text{Prob}(q_l \cap 0)
\]

where \( q_l = \sum_{i=1}^{n} \hat{\psi}_i^2 \). Additionally, when the null hypothesis \( H_0: q_1 = 0 \) held, the quadratic form \( q_l \) had the characteristic function \( \phi_l(t) = \prod_{i=1}^{n} (1 + 2(\lambda_l - c_0)t)^{-1} \). The distribution function was then uniquely determined by this characteristic function:

\[
F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^x e^{it\phi_l(-t)} - e^{it\phi_l(t)} dt.
\]

We then tested \( H_0: \phi_4 = 0 \) given \( \phi_1 = \phi_2 = \phi_3 = 0 \) against \( H_0: -\phi_4 \neq 0 \). To each autoregressive \( S. \text{ damnosum s.l. riverine larval habitat model using the marginal probability (p-value) and}

\[
F(0) = \frac{1}{2} + \frac{1}{2\pi} \int_0^x (\phi_4(-t) - \phi_4(t)) dt = \prod_{i=1}^{n} (1 - 2(\lambda_4 - \bar{d}_4)t)^{-\frac{1}{2}}
\]

and \( \bar{d}_4 \) which was the calculated value of the fourth-order DW statistic.

The DW statistics were then used to determine whether the OLS regression estimates indicated significant serial uncertainty correlation with an estimated order of a lagged covariance of 1 in the larval habitat models. The AUTOREG procedure corrected for the serial correlation using the YW method. The DW statistics indicated that uncertainty correlation was only slightly significant in the YW corrected models. The YW estimates for the first-order serial autocorrelation Bagan model indicated a \( R^2 = 0.574 \), \( F \) statistics of 37.159, and DW score of 3.877 while YW estimates for the Sarakaw-Kpleou model indicated a \( R^2 = 0.4911 \), \( F \) statistics of 38.541, and DW score of 3.713.

The distribution of the error residuals in the second-order autocovariance matrices was then assessed. Initially, the parameter estimators sampled in the riverine epidemiological study sites were qualitatively assessed using their cluster-specific repeated predictor covariate coefficient indicator measurement values (Table 2). Variance decomposition based upon pseudo-\( R^2 \) values and model diagnostics was then obtained for each model by regressing the observed on the predicted standardized residuals (Table 3). The maximum value of \( I \) was then obtained by all of the variation of \( z \) as explained by the eigenvector \( u_i \) which corresponded to the highest eigenvalues in the matrices in both models. Thereafter, \( \text{cor}^2(u_i, z) = 1 \) and \( \text{cor}^2(u_i, z) = 0 \) for \( i \neq 1 \) and the maximum value of \( I \) was deduced for

\[
I(x) = \sum_{i=1}^{n-1} I(u_i) \text{cor}^2(u_i, z)
\]

which was equal to \( I_{\text{max}} = \hat{\lambda}_1(\frac{n}{1-n}) \) in the model residuals. The minimum value of \( I \) in the autocovariate parameter error matrices was then obtained as the variation of \( z \) which in this research was explained by the eigenvector \( u_{n-r} \) corresponding to the lowest eigenvalue \( \lambda_{n-r} \). This minimum value was then determined to be equal to \( I_{\text{min}} = \hat{\lambda}_{n-r}(\frac{n}{1-n}) \). The autoregressive error matrices then
Table 2. *Simulium damnosum s.l.* parameter estimates with cluster specific repeated within cluster based covariate measures.

| Breeding site and sampled covariates | With no random effects | With random effects |
|--------------------------------------|------------------------|---------------------|
|                                      | Parameter estimate     | Standard error      |
| Bagan                                |                        |                     |
| Intercept                            | –1.5555                | 0.1417              |
| TURB                                 | 0.0126                 | 0.0032              |
| RCKS                                 | 0.01459                | 0.0127              |
| AQVEG                                | 0.2222                 | 0.0347              |
| HGVEG                                | 0.6090                 | 0.0774              |
| MMBR                                 | 0.3334                 | 0.0693              |
| Random effects                       | 0                      | 1                   |
| Scale                                | 3.4338                 | 2.9112              |
| Sarkawa-Kpleou                       |                        |                     |
| Intercept                            | –1.4404                | 0.1507              |
| DISHAB                               | 0.0133                 | 0.0032              |
| DDVEG                                | 0.0142                 | 0.0128              |
| HGVEG                                | 0.2230                 | 0.0347              |
| AQVEG                                | –0.1852                | 0.0864              |
| Random effects                       | 0                      | 1                   |
| scale                                | 3.4259                 | 3.3137              |

Table 3. Variance decomposition based upon pseudo-$R^2$ values and model error diagnostics obtained by regressing the observed on the predicted standardized rates for the Bagan and the Sarkawa-Kpleou study sites.

| Clustering for repeated measures | Bagan site | Sarkawa-Kpleou site |
|---------------------------------|------------|---------------------|
| Common covariates               | 0.3102     | 0.3098              |
| Clustering-specific covariates  | 0.0264     | 0.0308              |
| SURE                             | 0.1981     | 0.0282              |
| Negative spatial autocorrelation| 0.0172     | 0.0221              |
| Positive spatial autocorrelation| 0.0099     | 0.0041              |
| Pseudo-$R^2$                     | 0.5618     | 0.3950              |
| Spatial filter $R^2$             | 0.1469     | 0.4375              |
| Spatial filter MC                | –0.03267   | –0.32958            |
| P(S-W) for random effects        | –0.0001    | 0.4609              |
| P(S-W) for SURE                  | 0.0005     | 0.5475              |

revealed if the sampled predictor covariate coefficient estimates were not spatialized; the part of the variance explained by each eigenvector was equal to $\operatorname{cor}^2(u_i, z) = \frac{\lambda_i}{n} - 1$. Because the data in $z$ were randomly permuted in the riverine larval habitat models it was assumed that we would obtain this result. The set of $n!$ random permutations in the models, then revealed that: $E[I] = \frac{n}{\sqrt{n}} \sum_{i=1}^{n} \lambda_i = \frac{n}{\sqrt{n}} \text{trace}(\Omega)$.

Additionally, the riverine larval habitat model residuals demonstrated that $\text{trace}(\Omega) = \frac{\sqrt{n}}{n} I$ and that $E[I] = \frac{1}{\sqrt{n}}$ existed.

The diagonalization of the matrices generated from the sampled immature *S. damnosum s.l.* riverine habitat data also consisted of finding the normalized vectors $u_i$, which was stored as columns in the error matrices where, $U = [u_1 \cdots u_n]$, satisfied $\Omega = \text{HWH} = U \cdot U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$ and where $\Lambda = \text{diag}(\lambda_1 \cdots \lambda_n)$, $u_i^T u_i = ||u_i||^2 = 1$ when $u_i^T u_j = 0$ for $i \neq j$. The double centering of $\Omega$ implied that the eigenvectors $u_i$ rendered from the sampled residual predictor covariate coefficient estimates from the riverine larval habitat models were centered and that at least one eigenvalue was equal to zero. Introducing these eigenvectors in the original formulation of the Moran’s coefficient generated from the sampled data led to $I(x) = \frac{n}{\sqrt{n} \sqrt{\text{tr}\Omega}} \sum_{i=1}^{n} \lambda_i u_i u_i^T x = \frac{n}{\sqrt{n} \sqrt{\text{tr}\Omega}} \sum_{i=1}^{n} \lambda_i u_i u_i^T x = \frac{n}{\sqrt{n} \sqrt{\text{tr}\Omega}} \sum_{i=1}^{n} \lambda_i u_i u_i^T x$ in both models.

Marginal and conditional distributions from inverse-Wishart distributed matrices were then determined for spatially summarizing Bayesian inferences from the sampled riverine larval habitat data. In this research, $A \sim W^{-1}(\Psi, m)$ had an inverse-Wishart distribution. We partitioned the matrices $A$ and $\Psi$ conformably with each other using $A = [A_{11} \ A_{12}; A_{21} \ A_{22}]$, $\Psi = [\Psi_{11} \ \Psi_{12}; \Psi_{21} \ \Psi_{22}]$ where $A_{ij}$ and $\Psi_{ij}$ were $p_i \times p_j$ matrices. We then obtained $A_{11}$ which was independent of $A_{12}$ and $A_{22} \cdot 1$, where $A_{22} \cdot 1 = A_{22} - A_{21} A_{11}^{-1} A_{12}$, was the Schur complement of $A_{11}$ in $A$. Commonly, a finite element problem is split into nonoverlapping subdomains in an autoregressive cluster-based regression distribution models and the unknowns in the interiors of the subdomains are eliminated ($I$). In this research, the remaining Schur complement system on the unknowns associated with
subdomain interfaces was solved by the conjugate gradient method.

The model initially revealed that the conjugate gradient method was unstable with respect to the perturbations in the models. However, the conjugate gradient method we employed had an iterative method which provided monotonically improving approximations (i.e. \( x_k \)) which in this research was achieved using the required tolerance rates after a relatively small number of iterations was performed in both models. Our improvement was linear and its speed was quantified by the condition number \( \kappa(A) \) of the system matrix \( A \); where, the larger the \( \kappa(A) \) the slower the improvement in the residual foreground casted uncertainty estimates. Since some of our \( \kappa(A) \) were large, preconditioning was used to replace the original system \( Ax - b = 0 \) with \( M^{-1}(Ax - b) = 0 \) so that \( \kappa(M^{-1}A) \) got smaller than \( \kappa(A) \). In most cases, preconditioning is necessary to ensure fast convergence of the conjugate gradient method \( \delta \).

In this research, the preconditioned conjugate gradient method in the riverine larval habitat cluster-based autoregressive models took the form: \( x_0 := b - Ax_0 ; z_0 := M^{-1}r_0 ; p_0 := z_0 : k := 0 \) where: \( x_{k+1} := x_k + z_k p_k \sim r_{k+1} \sim r_k - x_k p_k ; \) if \( r_{k+1} \) was sufficiently smaller than the exit loop end and if: \( z_{k+1} := M^{-1}r_{k+1} + \beta_k z_k \), \( p_{k+1} := z_{k+1} + \beta_k p_k \) and \( k := k + 1 \). In both models the above formulation was equivalent for applying the conjugate gradient method without preconditioning; whereby, \( E^{-1} A(E^{-1})^T \hat{x} = E^{-1} b \) and where \( EET = M \) and \( \hat{x} = E^T x \).

The preconditioner matrix \( M \), was symmetric positive-definite and fixed, (i.e. stationary from iteration to iteration) in the models. We then compared the number of iterations with empirical distribution functions using the autoregressive residual variance estimates. In this research, the empirical distribution function was the autoregressive residual variance estimates. In this research, 160,000 Markov chain Monte Carlo (MCMC) replications were executed using the sampled data from each riverine epidemiological breeding study site. The first 10,000 were discarded as a burn-in set and the resulting 150,000 were weeded such that only the third replication result was retained.

The final MCMC data-set contained 50,000 replications. For quantifying ergodicity of the MCMC algorithm, assuming that all the samplers simultaneously satisfied the nested polynomial drift conditions, we determined that either when the number of nested drift conditions was greater than or equal to two, or when the number of drift conditions was one, the adaptive algorithm was ergodic. Ergodicity of an adative MCMC algorithm in a spatiotemporal vector insect larval habitat cluster-based regression model refers to positive recurrent aperiodic state of stochastic systems tending in probability to a limiting form that is independent of the initial conditions \( \theta \). For the Bagan epidemiological study site the diagnostic Shapiro-Wilk statistic had a null hypothesis probability of \( P(S - W) = 0.0025 \) and the random effects increased the pseudo-\( R^2 \) value to 0.8652 while the spatially structured random effects (SSRE) in the sampled data accounted for about 52% of the random effects in the model. For the Sarakawa-Kpleou study site the diagnostic Shapiro-Wilk statistic had a null hypothesis probability of \( P(S - W) = 0.1572 \) and the random effects increased the pseudo-\( R^2 \) value to 0.9973. The SSRE accounted for about 34% of the random effects in the model. The larval habitat map patterns in both study sites were characterized by overall negligible negative spatial autocorrelation for the high density ABR-stratified cluster; (the Moran’s coefficient was -0.0531 and -0.2240 for the Bagan and Sarakawa-Kpleou sites, respectively).

The final model output for the Bagan epidemiological study site detailing the sequential decomposition of the variance is shown in Table 5. The logistic regression model mean response was then estimated with quasi-likelihood techniques because of the presence of severe underdispersion (i.e. 0.0227) which also comprised the sampled predictor covariate coefficients estimates as revealed in Table 5. The final model output for the parameter estimators using: \( P(X \mid \Psi, m) = \int P(X \mid \Sigma) P(\Sigma \mid \Psi, m) d\Sigma = \frac{\prod_{j=1}^{m} \left[ \frac{1}{2\pi \sigma_j^2} \right]^{-1/2} \exp \left( -\frac{1}{2\sigma_j^2} \right) }{\prod_{j=1}^{m} \left[ \frac{1}{2\pi \sigma_j^2} \right]^{-1/2} \exp \left( -\frac{1}{2\sigma_j^2} \right) } \). The variance of the diagonal used the same formula in both models when \( i=j \), which was then simplified to: \( \text{var}(b_j) = \frac{z_i^2 \sigma_i^2}{(m-p-1)(m-p-3)} \). The mean was then \( E(B) = \frac{\Psi}{m-p-1} \). Thereafter, we calculated the variance of each element of \( B \) in the models which revealed \( \text{var}(b_j) = \frac{(m-p+1)z_i^2 + (m-p-1)\phi_j \phi_j}{(m-p)(m-p-1)(m-p-3)} \).

Other Bayesian specifications were then generated employing normal priors for each of the logistic regression-based coefficients. This solution had posterior mean regression coefficients and standard errors that were almost identical to those for a frequentist solution. In this research, 160,000 Markov chain Monte Carlo (MCMC) replications were executed using the sampled data from each riverine epidemiological breeding study site. The first 10,000 were discarded as a burn-in set and the resulting 150,000 were weeded such that only the third replication result was retained.
Sarakawa-Kpleou epidemiological study site also detailed the sequential decomposition of variance as shown in Table 6. Similarly as the Sarakawa-Kpleou study site, the logistic regression mean responses were estimated for the Bagan model using quasi-likelihood techniques because of the presence of severe underdispersion (i.e. deviance = 0.0017) which contained the predictor covariate coefficients estimates in Table 7.

We then constructed multiple simulation outputs from the validation model residuals. The simulation of each of our high and low ABR-stratified riverine larval habitat cluster-based regression parameter estimators was a two-step process. In Step 1, we simulated \( Y_i \sim N(c\sqrt{Z_i} + \beta X_i, 1/w_i) \) independently for \( i \). We displayed the results when \( c = 1 \) and, thus, the outcome was \( Y_i \) using the residual explanatory data from the initial autoregressive riverine larval habitat model outputs. We found that the power increased based on the positive relationship (i.e. \( \gamma > 0 \), \( \beta > 0 \)) between the sampled georeferenced predictor covariate coefficient estimates. This was expected in our models since \( E(Y_i|Z_i) = (c\sqrt{2} + \beta \gamma)Z_i \), directly depended on the values of \( \gamma \) and \( \beta \). We noticed that when we adjusted for \( X_i \), the power of the validation models decreased revealing a stronger association between the predicted residual autoregressive uncertainty estimates using \( X_i \) and \( Z_i (\gamma \rightarrow \infty) \). Our first simulation assessed the unadjusted analyses in the models when there was clustering of the residual within varying and constant predictor covariate coefficients effects which in this research was indirectly induced by \( X_i(E(Y_i|Z_i) = \beta \gamma Z_i) \). The individual-sampled riverine larval habitat level explanatory variable outcome values in each riverine epidemiological breeding study site was then quantified by \( V_i = \hat{U}_i + \hat{\beta}_i X_i^f \). There was a normal distribution in the estimating equations for \( e_{ij} \) and \( X_i^f \) which was a vector in the models when the sampled predictor covariate coefficients measurement indicator values of the georeferenced indicator variable was quantified from the initial model outputs.

We then used the distribution on \( U_i \) and \( e_{ij} \) based on \( U_i^{\text{ind}}(\beta R, \sigma_R^2) \) and \( e_i^{\text{ind}}(0, \sigma^2) \) for verifying the empirical estimates for \( U_i \). Thereafter, we incorporated the individual-sampled residual intra-cluster predictive serial error correlation values using \( B_i = \sum_{i=1}^n |Y_{ij} - \hat{\beta}_i X_{ij}^f - \hat{U}_i|/n \). The residual estimates were quantified based on the relationship between the sampled georeferenced riverine larval habitat locations in each epidemiological riverine study site and \( Y_{ij} \), given \( X_{ij} \) and \( E(B_i)=0 \) and \( \text{Var}(B_i) = (\sigma^2 + \sigma_R^2)/n \) which, in turn, was used to verify all estimated weights from the autoregressive models using the inverse of the variance of \( B_i \), \( w_i = n/(\sigma^2 + \sigma_R^2) \). In the models, for a given sampled georeferenced riverine larval habitat [i.e. \((x_1, x_2)\)], \( \sqrt{nZ_{\text{loc}}} \) was the weighted sum of residuals. In this research, if a cluster-based predictive autoregressive estimate occurred in areas with a higher intensity of an outcome this implied a larger value of \( Z_{\text{loc}}(x_1, x_2) \).

Unfortunately, the exact distribution of \( Z_{\text{loc}}(x_1, x_2) \) could not be solved analytically in the models so an asymptotic equivalent distribution was used to approximate the true distribution of the sampled parameter estimators. We then considered the following expressions for validating the estimates using \((x_1, x_2), Z_{\text{loc}}(x_1, x_2) \),

| Parameter | Estimate | Standard Error | Chi-square |
|-----------|----------|----------------|------------|
| Intercept | -1.1448  | 0.0236         | 2351.41    |
| TURB      | -0.0854  | 0.0020         | 1836.79    |
| RCKS      | 0.1521   | 0.0071         | 459.18     |
| latitude  | 0.0230   | 0.0015         | 249.57     |
| AQVEG     | 0.4141   | 0.0150         | 762.26     |
| HGVEG     | -0.2247  | 0.0201         | 125.05     |
| FLVEG     | 0.2477   | 0.0202         | 149.79     |
| SURE      | 4.3325   | 0.0521         | 6924.57    |
| E29 (−SA) | 0.6417   | 0.0448         | 241.59     |
| E37 (−SA) | -0.7354  | 0.0428         | 294.91     |
| E41 (−SA) | -0.4152  | 0.0414         | 100.01     |
| E47 (−SA) | 0.8233   | 0.0419         | 387.29     |
| E49 (−SA) | -0.7189  | 0.0424         | 288.19     |
| E53 (+ SA)| 0.4589   | 0.0425         | 117.54     |
| Scale     | 0.1563   |                |            |

Table 4. The final detailed sequential decomposition of variance for the Bagan study site.

| Variance component | Partial pseudo-\( R^2 \) |
|--------------------|-------------------------|
| four covariates    | 0.3066                  |
| latitude           | 0.0183                  |
| SURE               | 0.5414                  |
| negative spatial autocorrelation (− SA) | 0.0781 |
| positive spatial autocorrelation (+SA) | 0.0433 |
| Total              | 0.9876                  |

Table 5. The logistic regression model mean response, which was estimated with quasi-likelihood techniques for the Bagan riverine study site.
Table 6. The final detailed sequential decomposition of variance for the Sarakawa-Kpleou study site.

| Variance component                      | Partial pseudo-$R^2$ |
|----------------------------------------|----------------------|
| Two covariates                         | 0.2971               |
| SURE                                   | 0.4324               |
| negative spatial autocorrelation (-SA) | 0.1818               |
| positive spatial autocorrelation (+SA) | 0.0415               |
| Total                                  | 0.9173               |

Table 7. The logistic regression model mean response, which was estimated with quasi-likelihood techniques for Sarakawa-Kpleou study site.

| Parameter   | Estimate | Standard error | Chi-square |
|-------------|----------|----------------|------------|
| Intercept   | -1.3475  | 0.0074         | 25652.9    |
| $I_{daboc}$ | 3.9688   | 0.0871         | 2082.24    |
| $I_{single}$| -0.5725  | 0.0174         | 1082.14    |
| SURE        | 18.7079  | 0.2777         | 4539.44    |
| E_{13} (-SA)| 0.2559   | 0.0110         | 539.36     |
| E_{17} (-SA)| 0.3560   | 0.0109         | 1071.41    |
| Scale       | 0.0412   |                |            |

where $v(x_1, x_2|b) = -\sum_{i=1}^n W_i(x_1, x_2|b)\partial\mu/\partial\beta = -\sum_{i=1}^n W_i(x_1, x_2|b)X_i$. Our validation models revealed $I(\beta)=-\partial U/\partial\beta$ and $G_i$ ($i = 1, \ldots, n$) were independent displaying a mean of 0 and variance of 1. It thus followed that the asymptotic conditional distribution of the $Z_{loc}(x_1, x_2|b)$, given the observed riverine larval habitat residual outputs ($Y_i, X_i, s_i, r_i$) ($i = 1, \ldots, n$), were equivalent to the distribution of $Z_{loc}(x_1, x_2|b)$, assuming that the georeferenced larval habitat geolocation, ($s_i, r_i$) was independent of the outcome (i.e. $Y$). These results were obtained by qualitatively assessing and then quantifying the independence between the forecasted residual predictor covariate error coefficient estimates under the null hypothesis. The asymptotic results from both riverine larval habitat models allowed us to approximate the null distribution of $\hat{S}_{loc}(x_1, x_2|b)$ employing multiple $n$ realizations of $\hat{Z}_{loc}(x_1, x_2|b), (\hat{Z}_{loc}(x_1, x_2|b), \ldots, \hat{Z}_{N,loc}(x_1, x_2|b))$ and by repeatedly simulating independent samples of $(G_1, \ldots, G_n)$, while adjusting the autoregressive residual estimates using $(Y_i, X_i, s_i, r_i)$ ($i = 1, \ldots, n$). A finite vector of length $M$ of the explanatory estimates was also denoted by $b=(b_1, \ldots, b_M)$ in each validation model where each $b_m$ represented the size of the ABR-stratified clusters. Accordingly, we defined the validation test statistics using

$$S_{loc} = \sup_{x_1, x_2} \left[ \sup_{x_1, x_2} \hat{Z}_{loc}(x_1, x_2|b_1), \ldots, \sup_{x_1, x_2} \hat{Z}_{loc}(x_1, x_2|b_M) \right]$$

which was conditional on the sampled estimates using

$$\hat{S}_{loc} = \sup_{x_1, x_2} \left[ \sup_{x_1, x_2} \hat{Z}_{loc}(x_1, x_2|b_1), \ldots, \sup_{x_1, x_2} \hat{Z}_{loc}(x_1, x_2|b_M) \right].$$

Thereafter, the empirical $p$-values were computed as

$$p - value = \frac{\sum_{i=1}^n I(z_{loc, j} < \hat{S}_{loc})}{s},$$

for each riverine larval habitat cluster-based regression model, where $\hat{S}_{loc}$ was the $\hat{S}_{loc}$ at the $j$th realization of $\hat{Z}_{loc}$. The residual model outputs were $\{x_1, x_2, b\}: Z_{loc}(x_1, x_2|b) \geq \hat{S}_{(95\%)}$, where $S_{(95\%)}$ was the 95th percentile for all $\hat{Z}_{loc}$ rendered from the forecasted estimates. In the riverine larval habitat models, we noted that when $E(Y|Z_i, \beta) = c\sqrt{Z_i^2 + \beta X_i}$ and $\text{Var}(Y) = 1$, $w_i$ was 1 for all $i = 1, \ldots, n$. Additionally, $Z_i$ was an important indicator value of the sampled residual predictor error covariate coefficients in both models if $i$ was within $Z_i$ and $X_i$ and had a varying $\beta$ (i.e. dependence of $Y_i$ on $X_i$) and $\gamma$ (i.e. dependence of $X_i$ on $Z_i$). We varied $\beta$ employing 2 to 2.5 by sequential 1 by 0.01, and $\gamma = 0.5$ or 1.0 there was no power to detect any sampled explanatory estimates in either model. However, when we allowed $\beta \geq 0$ and $\gamma = 0.5$ or 1.0 there was power to detect a prolific riverine larval habitat based on spatiotemporal field-sampled autoregressive forecasted count data in each riverine epidemiological study site was within normal statistical thresholds, but it did increase as expected with more positive dependence between $Z_i$ and $X_i$ ($\gamma > 0$, $\gamma \to \infty$) and stronger positive association between $X_i$ and $Y_i$ ($\beta > 0$, $\beta \to \infty$). Our models revealed that when $\beta \leq 0$ and $\gamma = 0.5$ or 1.0 there was no power to detect any sampled explanatory estimates in either model. However, when we allowed $\beta \leq 0$ and $\gamma = 0.5$ or 1.0 there was power to detect a prolific riverine larval habitat based on spatiotemporal field-sampled autoregressive forecasted count data in each riverine epidemiological study site was within normal statistical thresholds, but it did increase as expected with more positive dependence between $Z_i$ and $X_i$ ($\gamma > 0$, $\gamma \to \infty$) and stronger positive association between $X_i$ and $Y_i$ ($\beta > 0$, $\beta \to \infty$). Our models were conditional on $X_i$ as there was independence between $Y_i$ and $Z_i$. The autoregressive residual uncertainty outputs in both $S. damnosum s.l.$ riverine models revealed that the predictive power was equal to the Type I error rate of 0.05.

4. Discussion

In conclusion, varying coefficient cluster-based regression residuals, diagnostic Shapiro-Wilks statistics, respecified Bayesian priors and QuickBird visible and NIR data determined latent negative spatial error autocorrelation components in the residual predictive autoregressive intra-cluster $S. damnosum s.l$. larval habitat predictor covariate correlation analyses for both the Bagan and Sarkawa-Kpleou riverine epidemiological study sites. Designing and developing $S. damnosum s.l.$ riverine larval habitat management strategies in ArcGIS based on spatial statistical algorithms in SAS/GIS and PROC MCMC using sub-meter resolution satellite data and robust diagnostic residual intra-cluster predictor covariate error correlation estimates can provide an effective entomological tool to reduce prolific $S. damnosum s.l$. larval habitats based on spatiotemporal field-sampled count data in riverine ecosystems.
Notes on contributor
Benjamin G. Jacob received a PhD in environmental epidemiology from the School of Medicine at the University of Miami, FL, USA. Thereafter, he completed a National Institute of Health (NIH) postdoctoral fellowship in medical entomology at the University of Illinois at Urbana/Champaign in, IL, USA. Currently, he is a research assistant professor at the College of Public Health, Department of Global Health University of South Florida, Tampa, FL, USA.

References
(1) Toe, L.; Merriweather, A.; Unnasch, T.R. DNA Probe-based Classification of Simulium damnosum s.l.-Borne and Human-Derived Filarial Parasites in the Onchocerciasis Control Program Area. Am. J. Trop. Med. Hyg. 1994, 51 (3), 676–683.
(2) Crosskey, R.W. A Taxonomic Study of the Larvae of West African Simuliidae (Diptera: Nematocera) with Comments on the Morphology of the Larval Blackfly Head. Bull. British Museum (Natural History) Entomology 1960, 10 (6), 1–74.
(3) Paugy, D.; Fermon, Y.; Abban, K.E.; Diop, E.; Traoré, K. Onchocerciasis Control Programme in West Africa: A 20-year Monitoring of Fish Assemblages. Aquat. Living Resour. 1999, 12 (7), 363–378.
(4) Toe, L.; Tang, J.; Back, C.; Katholi, C.R.; Unnasch, T.R. Vector-Parasite Transmission Complexes for Onchocerciasis in West Africa. Lancet 1997, 349, 163–166.
(5) Garms, R. The Reinvasion of the Onchocerciasis Control Programme Area in the Volta River Basin by Simulium damnosum s.l. the Involvement of the Different Cytoespecies and Epidemiological Implications. Ann. Soc. Belgium Medicine Tropic. 1979, 61 (2), 193–198.
(6) Boatin, B.; Molyneux, D.H.; Hougard, J.M.; Christensen, O.W.; Alley, E.S.; Yaméogo, L.; Seketeli, A.; Dadzie, K.Y. Patterns of epidemiology and control of onchocerciasis in West Africa. J. Helminthol. 1997, 71 (7), 91–101.
(7) Gu, W.; Novak, R.J. Habitat-Based Modeling of Impacts of Mosquito Larval Interventions on Entomological Inoculation Rates, Incidence, and Prevalence of Malaria. Am. J. Trop. Med. Hyg. 2005, 73 (17), 5460–5552.
(8) Griffith, D.A. Spatial Autocorrelation on Spatial Filtering; Springer: Berlin, 2003.
(9) McCulloch, C.E.; Searle, S.R. Generalized, Linear and Mixed Models; Wiley-Interscience: New York, NY, 2005.
(10) Jacob, B.G.; Griffith, D.A.; Muturi, E.J.; Caamano, E.X.; Githure, J.I.; Novak, R.J. A heteroskedastic error covariance matrix estimator using a first-order conditional auto-regressive Markov simulation for deriving asymptotical efficient estimates from ecological sampled Anopheles arabiensis aquatic habitat covariates. Mal J. 2009, 28 (2), 216–225.
(11) Cressie N. Aggregation in Geostatistical Problems. In Geostatistics troia 92: Soares, A., Ed.; Kluwer: Dordrecht, 1993; pp. 25–36.