Journal of Statistics: Advances in Theory and Applications
Volume 10, Number 1, 2013, Pages 1-136

FINITE-DIFFERENCE DERIVATIVES OF A FIRST-ORDER INTEGRAL APPROXIMATION QUANTIZED WITH A DEFAULT DUAL QUASI-NEWTON OPTIMIZER AND A PSEUDO-LIPSCHITZIAN PROPERTY FOR PREDICTIVE MAPPING SPATIALLY INHOMOGENEOUS *Similium damnosum s.l.* EXPLANATORY COVARIATES

BENJAMIN G. JACOB[1], ROBERT J. NOVAK[1], LAURENT TOE[2], MOUSSAS S. SANFO[2], SEMIHA CALISKAN[1], ROSE TINGUERIA[4], ALAIN PARE[3], LAURENT YAMEOGO[3], DANIEL GRIFFITH[5] and THOMAS R. UNNASCH[1]

[1]Department of Global Health
College of Public Health
University of South Florida
Tampa, FL
USA
e-mail: bjacob1@health.usf.edu

[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

2010 Mathematics Subject Classification: 62H11.

Keywords and phrases: PROC NL MIXED, Moran's coefficient, SAS Macro Win BUGSio, Bayesian, onchocerciasis.

Received July 1, 2013

[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 at Dallas
800 West Campbell Road
Richardson, TX 75080-3021
USA

Abstract

Currently the most common hypothesis-testing methods for predictive seasonal risk modeling arthropod-related infectious disease data is employing traditional multivariate analogues. Unfortunately, statistical programs are not stringent enough for seasonally quantitating error and error assumptions in the probability plots of regressed forecasts for accurately interpolating endemic transmission regions in an epidemiological study site. In this paper, we propose a Gaussian process in a spatial filter analyses and a Bayesian matrix for deriving qualitative probabilistic inferences from an ecological regressed dataset of noisy georeferenced riverine-based larval habitats of *Similium damnosum*, a black fly vector of onchocerciasis. Our intention was to simulate optimally unbiased seasonal endemic transmission-oriented explanatory covariate coefficients based on spatial aggregations of productive habitats within a riverine epidemiological study site by introducing a latent variable within a non-linear autoregressive equation. Autocorrelation scatterplots revealed that the Moran's coefficient was 0.067, while the Gearys ratio was 0.891 in the forecasts. Improvement of fit of a WinBUGS hierarchical Bayesian model then revealed that the adjusted covariate Presence of hanging vegetation was statistically important to prolific sampled habitats.

1. Introduction

Quantitative autoregressive predictive seasonal vector arthropod-related risk modeling is one of the most challenge areas presently in ecology. One of the reasons for these challenges is that vector ecologists and local abatement district managers commonly regress large ecological datasets of highly correlated multivariate vector arthropod-related biotic and abiotic endemic transmission-oriented explanatory covariate coefficients jointly in one single regression-based matrix. As such,

standard estimators like the unstructured maximum likelihood (ML) estimator or the restricted maximum likelihood (REML) estimator in such programs as SAS can be very unstable for deriving optimal model residual forecasts. In statistics, the restricted (or residual, or reduced) ML (REML) approach is a particular form of ML estimation which does not base estimates on an ML fit of information, but instead uses a likelihood function calculated from a transformed set of data so that nuisance parameters have no effect [1]. A nuisance parameter in a multivariate predictive seasonal vector arthropod-related risk model is any parameter which is not of immediate interest, but which must be accounted for in the analysis of those sampled estimators which are of interest (e.g., the variance σ^2 of a normally regressed seasonal forecasted vector arthropod-related georeferenced distribution where the mean, μ, is of primary interest) [2].

In the case of variance component estimation of an ecological dataset of regressed multivariate seasonal-sampled vector arthropod-related endemic transmission oriented explanatory covariate coefficients, the original dataset may be replaced by a set of contrasts calculated from the likelihood function. By so doing, a probability distribution of the seasonal multivariate vector arthropod-related regressed contrasts can be parsimoniously and accurately extrapolated. For example, motivated by the fact that the method of least-squares is presently one of the leading principles in parameter estimation in seasonal multivariate vector arthropod-related predictive risk model construction, a vector ecologist and/or a local abatement district manager may develop a method of least-squares variance component estimation (LS-VCE) for statistically evaluating seasonal-sampled endemic transmission-related parameters. Thereafter, LS-VCE can provide a unified least-squares framework for estimating the unknown parameters in the residual forecasts of functional and stochastic models. Fortunately, LS-VCE has a similar insightful geometric interpretation as standard least-squares. As such, properties of the normal equations, estimability, orthogonal projectors, precision of estimators and non-linearity can be easily established for any empirical seasonal sampled multivariate vector arthropod-related

ecological dataset. Additionally, measures of inconsistency such as the quadratic form of residuals and the Wilcoxon Rank Sum (w)-test statistic can be generated in any current commercial statistical package (e.g., SAS Proc Univariate and StatXact) employing the sampled dataset, which would then logically lead to applying hypotheses testing for constructing robust stochastic multivariate endemic transmission oriented risk-based interpolators.

The w test statistic is the sum of the ranks from population X [1]. Interestingly, the w test can be used to test the null hypothesis that two seasonally-sampled multivariate vector arthropod-related populations X and Y have the same continuous distribution. For example, a vector ecologist and/or a local abatement district manager may choose to regress an empirical ecological dataset of independent random samples $x1, x2, \ldots, xm$ and $y1, y2, \ldots, yn$, of sizes m and n, respectively, from two seasonal-sampled epidemiological study site vector arthropod-related larval habitat populations. After merging the sampled data in VARCOMP each seasonal measurement then be hierarchically ranked. All sequences of ties can then be assigned an average rank. Assuming that the two populations have the same continuous distribution, then W would have a mean and standard deviation given by $\mu = m(m + n + 1)/2$ and $s = \text{Sqrt}[m\,n(N + 1)/12]$, where $N = m + n$ could optionally be used for testing the null hypothesis [i.e., Ho: there is no difference in distributions]. Conversely, a one-sided alternative hypothesis would be that the first population yields lower seasonal multivariate vector arthropod-related risk measurements. This alternative may be employed to determine if W is unusually lower than its expected value μ in the residual forecasts targeting the seasonal endemic transmission-oriented explanatory covariate coefficients of statistical significance. In this case, the p-value may be given by a normal approximation. Thereafter, by letting $N \sim N(\mu, s)$ and computing the left-tail $P(N <= W)$ using continuity correction for seasonally quantitating parameter significance levels, misspecifications may be identified. For example, W may be determined to be misspecified (e.g., regressed larval habitat value much

higher than its expected value) in the residually forecasted derivatives. Thus, the alternative hypothesis would yield more precise measurements from the regressed seasonal vector arthropod-related endemic transmission oriented explanatory covariate coefficients.

Thereafter, the p-value for a robust seasonal predictive vector arthropod-related multivariate endemic transmission-oriented risk model may be provided by the right-tail $P(N >= W)$ using continuity correction. If the two sums of ranks from each sampled larval habitat population, for example, are close, then a two-sided alternative may be employed (e.g., Ha: there is a difference in distributions). In such a case, the p-value may be given by twice the smallest tail value $(2 * P(N <= W)$ if, $W < \mu$, or $2 * P(N >= W)$ if, $W > \mu)$ in the risk model residual forecasts for accurately statistically targeting the statistically significant vector multivariate arthropod-related georefernced endemic transmission-oriented explanatory covariate coefficients in each distribution.

Thereafter, specific procedures can be employed to construct other statistical models for regressing the residual vector multivariate arthropod-related endemic transmission-oriented explanatory covariate coefficients. For example, the variance components procedure in SAS can be employed to construct a robust mixed-effects seasonal predictive risk models, whereby estimates of the contribution of each random effect to the variance of the dependent variable (e.g., total larval density count values) may be efficiently quantitated. This procedure may be particularly interesting for conducting analysis on mixed seasonal multivariate predictive vector arthropod-related endemic transmission models, such as split plots, univariate repeated measures, and random block designs. By calculating variance components using multiple model outputs, a vector ecologist and/or a local abatement district manager may then determine where to focus attention more efficiently (e.g., aggregation of prolific larval habitats based on spatiotemporal field-sampled count data), in order to reduce the variance in the residually forecasted derivatives for optimally targeting unbiased seasonal endemic transmission-oriented explanatory regressors.

Four different methods are currently available for estimating the variance components: minimum norm quadratic unbiased estimator (MINQUE), analysis of variance (ANOVA), ML, and REML. These methods are available in a number of general-purpose statistical software packages, including Genstat (the REML directive), SPSS (the MIXED command), Stata (the xtmixed command), and R (the lme4 and older nlme packages), as well as in more specialist packages, such as MLwiN, HLM, ASReml, and CropStat. Default output for these programs would also include variance component estimates for residually forecasting robust derivatives delineating the statistically significant endemic transmission-oriented explanatory covariates. If the ML method or the REML method is employed for constructing a seasonal predictive multivariate vector arthropod-related endemic transmission-oriented model, an asymptotic covariance matrix table may be also displayed. Other available outputs for regressed spatiotemporal vector arthropod-related ecological datasets includes an ANOVA table and expected mean squares for the ANOVA method and an iteration history for the ML and REML methods.

Fortunately, estimation variance components procedures in many statistical programs are fully compatible with the GLM univariate procedures. For example, the univariate procedure in PROC GLM would provide regression analysis and analysis of variance for one dependent variable (e.g., total seasonal sampled larval density count) using one or more factors and/or predictor variables when constructing a robust predictive multivariate vector arthropod-related risk model. The factor variables would then divide the sampled larval habitat population, for example, into various smaller linear regression-based groups. Using this GLM procedure, a vector ecologist or local abatement district manager may then test null hypotheses about the effects of multiple predictor variables on the means of various groupings of a single dependent variable. Specific interactions (e.g., random) in the risk model between factors as well as the effects of individual factors may be then efficiently quantitated. In addition, the effects of the sampled explanatory covariate coefficients and their interactions can be determined. For a robust regression analysis, the independent seasonal sampled vector arthropod-

related predictor variables can thereafter be specified as covariates in the model. Both balanced and unbalanced model outputs can also be tested to determine optimal estimators. A design is balanced if each cell in the model contains the same number of cases [1].

In addition to testing hypotheses, PROC GLM would produce estimates of the sampled seasonal vector arthropod-related endemic transmission-oriented parameters. The program would do so by constructing a regression model which would relate Y to a function of X and β in the model where the unknown parameters would be denoted as β, the independent variables as x and the dependent variable would be Y. The approximation could then be formalized as $E(Y|X) = f(X, \beta)$.

To carry out the analysis robustly however, the form of the function f must be specified in the vector arthropod-related model. The form of this function would be based on knowledge about the relationship between Y and X that does not rely on the seasonal sampled vector arthropod-related data. If no such knowledge is available, a flexible or convenient form for f can be alternatively chosen.

Assuming that the vector of unknown parameters β is of length k in an empirical sampled dataset of seasonal vector arthropod-related parameter estimators datasets, a regression analysis can also then provide information about the dependent variable Y (e.g., total spatiotemporal larval density count data) to specific endemic transmission oriented covariates (e.g., weekly rainfall).

If N sampled vector arthropod-related data points of the form (Y, X) are observed, where $N < k$, most classical approaches to multivariate seasonal vector arthropod-related regression analysis cannot be performed since the system of equations defining the regression model would be underdetermined due to insufficient data to recover β. If exactly $N < k$ vector arthropod-related seasonal sampled data points are observed in PROC GLM, for example, and the function f is linear, the equations $Y = f(X, \beta)$ can be solved exactly rather than approximately. This would reduce to solving a set of N equations with N unknowns (i.e., the

elements of β), which would then render a unique solution as long as the X are linearly independent in the risk model. If f is nonlinear in the model, a solution may not exist or conversely many solutions may exist. The most common situation in predictive risk mapping vector arthropod seasonal sampled endemic transmission oriented explanatory covariate coefficients is where $N > k$ data points are observed. In this case, there is enough information in the ecological sampled datasets to estimate a unique value for β that best fits the multivariate vector arthropod-related seasonal sampled datasets. Thereafter, the regression analysis can provide the tools for finding a solution for unknown parameters (e.g., β) that will, for example, minimize the habitat distance between the measured and predicted larval density count values of the dependent variable Y. Interestingly, under certain statistical assumptions, the regression analysis in PROC GLM would use the surplus seasonal sampled vector arthropod-related endemic transmission oriented data to provide statistical information about the unknown parameters β and predicted values of the dependent variable Y as well.

Thereafter, priori contrasts would be available to perform hypothesis testing using any dataset of empirical sampled vector athropod-related explanatory covariates. By so doing, an overall F test may be employed to quantitate exact parameter estimator significance. Post hoc tests may also be employed to evaluate differences among specific means. Estimated marginal means can also be generated to provide estimates of predicted mean values for the cells in seasonal-sampled vector arthropod related endemic transmission-oriented risk model profile plots (i.e., interaction plots) based on calculated means which can allow further visualization of the sampled estimator's relationships to a dependent variable in parameter space. Residuals, predicted values, Cook's distance, and leverage values can then be saved as new variables in the data file for checking seasonal assumptions. By doing so, weighted least square (WLS) would allow specification of any sampled vector arthropod-related predictor variable employed during the construction of the risk model. Thereafter, to compensate for variations in the precision of the georeferenced vector multivariate arthropod-related seasonal-sampled

measurements in the empirical ecological dataset of explanatory covariate coefficients, a robust uncertainty estimator can be employed for attaining optimal residual forecasts.

Alternatively, REML may be employed as a method for fitting linear mixed robust predictive seasonal multivariate vector arthropod-related endemic transmission-oriented risk models in PROC GLM. In contrast to a simple ML estimation, REML may produce unbiased estimates of variance and covariance parameters in risk model residual forecasts for targeting endemic transmission oriented explanatory covariate coefficients. By doing so, a standard approach may be also employed in PROC GLM to further stabilize a regression-based matrix. Computing the REML estimator under some simple structure would only then require estimation of a few parameters for determining compound symmetry or independence in the residual forecasts. However, these estimators may not be consistent unless the hypothesized structure is correct. If the interest of the vector ecologist and/or a local abatement district manager focuses solely on estimation of only some specific vector arthropod-related regression coefficients, for example, with correlated or longitudinal data, a sandwich estimator of the covariance matrix may be alternatively employed to provide standard errors for the estimated coefficients.

Sandwich estimators for standard errors are often useful when model based estimators are very complex and difficult to compute and robust alternatives are required [1]. Consider the fixed part of a seasonal vector arthropod-related multivariate predictive risk model (e.g., $\hat{\beta} = (X^r \hat{V}^{-1} X)^{-1} X^r \hat{V}^{-1} Y$) where covariance matrix is given by $(X^r \hat{V}^{-1} X)^{-1} X^r \hat{V}^{-1} \operatorname{cov}(Y|X) \hat{V}^{-1} X (X^r \hat{V}^{-1} X)^{-1}$. If the vector ecologist or a local abatement district manager replaces the central covariance term by the usual (i.e., normal) model based value, -V, the formula $(X^r \hat{V}^{-1} X)^{-1}$ then could be obtained with sample estimates being substituted liberally. The sandwich estimator would then be formed in the predictive multivariate vector arthropod-related endemic transmission model by replacing the estimate of the central covariance term, $\operatorname{cov}(Y|X)$, by an empirical

estimator based on the block diagonal structure cross product matrix, namely, $V^* = \widetilde{Y}\widetilde{Y}^r$, $\widetilde{Y} = Y - \hat{X}\beta$, $E(\widetilde{Y}\widetilde{Y}^r) = V$. For the estimated set of residuals for the j-th block at level h, the sub/superscript h would be then given by $\hat{u} = \Omega Z^r_{(j)} V^{-1} \widetilde{Y}$. Thereafter, to obtain consistent estimators of the covariance matrix of these residuals, comparative or diagnostic estimators may be chosen. By doing so, the diagnostic estimator would then be given by $E(\hat{u}\hat{u}^r) = \Omega Z^r_{(j)} V^{-1} V^* V^{-1} Z_{(j)} \Omega$. Interestingly, if the predictive multivariate seasonal endemic transmission-oriented risk model based estimator $V' = V$ is employed for parameter estimator significance testing,, this estimator would reduce to the expression of the cross product matrix estimator V'. A comparative seasonal predictive multivariate vector arthropod-related risk model estimator then could be expressed as $E[(\hat{u} - u)(\hat{u} - u)^r] = \Omega Z^r_{(j)} V^{-1} V^* V^{-1} Z_{(j)} \Omega + \Omega - 2\Omega Z^r_{(j)} V^{-1} Z_{(j)} \Omega$ when the model based estimator is $= V$.

Fortunately, there exists two general shrinkage approaches for estimating the covariance matrix and regression coefficients in most commercial statistical packages for robust vector arthropod-related endemic transmission oriented residual uncertainty modeling. The first involves shrinking the eigenvalues of the unstructured REML estimator. The second involves shrinking an unstructured estimator toward a structured estimator. For both cases, the multivariate seasonal-sampled vector arthropod-related data parameter estimator quantitation methods would calculate the amount of shrinkage. These estimators may then be determined to be consistent and thus provide asymptotically efficient estimates for the seasonal-sampled vector arthropod-related regressed coefficients. Simulations have shown improved operating characteristics of the shrinkage estimators of the covariance matrix and the regression coefficients in finite samples [1]. The final estimator chosen would then optimally include a combination of both shrinkage approaches in the multivariate vector arthropod-related residual forecasts (i.e., shrinking the eigenvalues and then shrinking toward structure). By doing so, constructive recommendations can be provided for making robust inferences employing a particular shrinkage estimator that could then

provide a reasonable compromise between structured and unstructured seasonal-sampled vector arthropod-related parameter estimators. This is of course assuming that the coefficients remain consistent under misspecifications of the covariance structure. With large matrices derived from seasonal sampled vector arthropod related empirical datasets, the efficiency of the sandwich estimator would however become rather worrisome since the independent variables would not be altogether perfectly correlated. Instead the estimates and their standard errors would be high. For example, the estimated beta weights may be larger than 1.0 indicating multicollinearity exists in the multivariate vector arthropod-related residual forecasts.

Multicollinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the other(s) with a nontrivial degree of accuracy. In this situation, the seasonal vector arthropod-related regression coefficient estimates may change erratically in response to small changes in the model. Multicollinearity in a seasonal multivariate vector arthropod-related endemic transmission-oriented risk model would not reduce the predictive power or reliability of the predictive risk model as a whole, at least within the sample covariate coefficient themselves, but it would affect calculations regarding individual observational predictors. That is a regression-based predictive multivariate seasonal vector arthropod-related risk model with correlated endemic transmission-oriented predictors may indicate how well the entire bundle of regressed explanatory covariate coefficients forecasts the outcome variable, but it may not give valid results about any individual sampled predictor.

Note that multicollinearity is a multivariate problem in a seasonal predictive vector arthropod-related risk model not a bivariate problem [2]. That means that a simple perusal of the bivariate correlation matrix may not be sufficient to eliminate consideration of the problem of multicollinearity in the risk model. The problem is not only that the independent variables may be highly correlated, but that only one independent variable may only be highly correlated in the model. Thus, vector ecologist and/or a local abatement district manager would have to

examine the R^2 of each independent variable regressed against the independent variables. Fortunately, this is easy with most statistical programs (e.g., SPSS) since multicollinearity diagnostics are commonly rendered with the model output values. For example, the tolerance printed on any SPSS output would be the proportion of variance in any seasonal sampled vector arthropod-related independent variable (e.g., total seasonal larval density count) which is not explained by its relationship with the other independent variables. As such, the final forecasts values would be better expressed as $1 - R^2$, with the R^2 resulting from regressing a particular seasonal-sampled vector arthropod-related independent variable (e.g., total weekly rainfall) against all other independent variables.

Fortunately, there are many clues that can be provided by the covariance matrix in most commercial statistical software packages about the robustness of the predictive regression coefficients. For example, if the regression coefficient values are "large", this would indicate a problem of multicollinearity in the residual forecasts targeting the statistically significant multivariate vector arthropod-related endemic transmission-oriented explanatory covariates. As a rule of thumb, if any of the variance inflation factor (VIF) in a predictive multivariate seasonal vector arthropod-related risk model is greater than 10, there is a multicollinearity problem. In statistics, the VIF quantifies multicollinearity in an ordinary least squares (OLS) regression analysis by providing an index that measures how much the variance (i.e., the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity [1]. The OLS a method for estimating the unknown parameters which minimizes the sum of squared vertical distances between the observed responses and the responses predicted by the linear approximation approximation [2]. Then, if the following linear model with k independent variables: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$ is considered for fitting an empirical sampled seasonal dataset of vector arthropod-related endemic transmission oriented explanatory covariate coefficients, the standard error of the estimate of β_j would be the square root of the $j + 1$.

Further, $j+1$ element of $s^2(XX)^{-1}$, where s is the root mean squared error (RMSE), may then be calculated when constructing a seasonal vector arthropod-related endemic transmission-oriented uncertainty risk model. The RMSD is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed [1]. The RMSD serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power in a seasonal predictive vector arthropod-related model. For example, Jacob et al. [2] determined that RMSE[2] is an unbiased estimator of the true variance of the error term, σ^2, in a seasonal black fly (i.e., *Similium damnosum s.l.*) riverine larval habitat model employing multiple georeferenced field and remote sampled spatiotemporal parameter estimators. The authors constructed the regression model for determining robust predictors of onchocerciasis in a riverine epidemiological study site in Togo. Onchocerciasis is a parasitic disease transmitted to humans through the bite of a black fly of the genus *Similium*, which causes impaired lymphatics, eyes lesions, and blindness. In their model, X was the regression design matrix (i.e., a matrix such that $X_{i,\,j+1}$ was the sampled value of the j-th independent variable for the i-th *S. damnosum s.l.* larval habitat-related seasonal-sampled observation whereby $X_{i,1}$ equaled 1 for all i). The authors then found that the square of this standard error in the model estimated the variance of the estimate of β_j, which was expressed by using $\widehat{\operatorname{var}}\left(\hat{\beta}_j\right) = \frac{s^2}{(n-1)\,\widehat{\operatorname{var}}\left(X_j\right)} \cdot \frac{1}{1-R_j^2}$, where R_j^2 was the multiple R^2 for the regression of X_j against the other sampled covariate coefficients [i.e., a regression that did not involve the total count response variable Y]. This identity separated the influences of several distinct factors on the variance of the coefficient estimate since s^2 revealed a greater scatter in the seasonal sampled *S. damnosum s.l.* larval habitat data around the regression surface which then subsequently lead to proportionately more variance in the coefficient estimates.

Further, the n seasonal sample size (i.e., 31 larval habitats) resulted in proportionately less variance in the coefficient estimates while $\widehat{\text{var}}(X_j)$ revealed a greater variability in the sampled endemic transmission oriented explanatory covariate Distance from a georeferenced larval habitat capture point, which then lead to proportionately less variance in the corresponding forecasted coefficient estimate. The authors then determined that the remaining term $1/(1-R_j^2)$ was the VIF which reflected all other factors that influenced the uncertainty in the coefficient estimates. Interestingly, the authors noted that the VIF equaled 1 when the vector X_j was orthogonal to each column of the design matrix during the regression of X_j based on the other sampled larval habitat covariate coefficient values. By contrast, the VIF was greater than 1 when the vector X_j was not orthogonal to all columns of the design matrix during the regression of X_j based on the other coefficient values. Finally, the VIF was found to be invariant to the scaling of the seasonal-sampled predictor variables. That is, the authors had the ability scale each seasonal-sampled *S. damnosum s.l.* larval habitat predictor variable X_j by a constant c_j without changing the VIF. By estimating the regression equations, the riverine larval habitat model revealed that the correlations among the independent variables was 813 thus, multicollinear variables were present in the residually forecasted derivatives.

Frameworks for coding protocol in some statistical packages (e.g., R, Stata) may help analyze hierarchical predictive linear vector arthropod-related error-based model applications. These issues could include: (a) model development and specification, which would cover issues of centering, selection of predictors, specification of covariance structure, fit indices, generalizability, and checks on specification; (b) data considerations including distributional assumptions, outliers, measurement error for predictors and outcomes, power, and missing data; (c) estimation procedures including alternative procedures such as bootstrapping; and (d) hypothesis testing for making statistical inferences about variance parameters and fixed effects. For this purpose, previously two major

methods: expectation-maximization (EM) REML and Bayesian via Gibbs sampling have been combined to determine unbiased seasonal multivariate vector arthropod-related parameter estimators.

Expectation-maximization REML is regarded as a very stable algorithm that is able to converge when covariance matrices are close to singular, however, it is slow. Further, convergence problems can occur with random regression models, especially, if the starting vector arthropod seasonal sampled endemic transmission-oriented risk-based covariate coefficient values, for example, are much lower than those at convergence. Average information (AI) REML is much faster for common problems, but it relies on heuristics for convergence. Regardless, it may be very slow or even diverge when regressing complex seasonal predictive multivariate vector arthropod-related model parameter estimators. REML algorithms for general models become unstable with larger number of traits [1]. Conversely, REML by canonical transformation is stable for regressing smaller sampled datasets but, in such cases only a limited class of models can be supported in current software platforms.

In general, REML algorithms are difficult to program. Bayesian methods via Gibbs sampling are much easier to program with REML, especially for complex models such as predictive seasonal multivariate vector arthropod-related endemic transmission-oriented risk models as they can support much larger datasets; however, the termination criterion can be hard to determine. Computing speed will then also vary with computational optimization techniques causing some large vector arthropod-related endemic transmission-oriented parameter estimator datasets and complex models to be unsupported in a reasonable time.

Parameter expanded and standard EM algorithms may be also described for reduced rank estimation of covariance matrices constructed from seasonal vector arthropod-related explanatory covariate coefficients by REML, but for model fitting only. By doing so, leading principal components can be then quantized in a robust seasonal predictive multivariate vector arthropod-related endemic transmission-oriented model. Convergence behaviour of these algorithms may then be examined. Implications for practical regression-based seasonal vector arthropod-related predictive risk-based analyses can then be seasonally

summarized. It may be shown, for example, that EM type algorithms are readily adapted to reduced rank estimation and convergence when constructing a robust seasonal predictive vector arthropod-related endemic transmission oriented risk model. However, as is well known for the full rank case, the convergence in the model would be linear and thus slow. Hence, these algorithms may be only useful in combination with the quadratically convergent average information algorithms, in particular, in the initial stages of an iterative solution scheme for predictive multivariate risk mapping seasonal sampled vector arthropod-related risk-based georeferenced feature attributes. Unfortunately, estimation based on the standard assumptions may still lead to inconsistent forecasts for targeting the statistically significant endemic transmission-oriented estimators as the residual outputs may not reflect the true correlation coefficient values in the ecological empirical sampled dataset.

Complications can also arise when constructing univariate statistics and predictive regression-based models (e.g., Poisson, Logistic) based on seasonal-sampled empirical arthropod-related explanatory covariate coefficients. For example, commonly seasonal-sampled vector arthropod-related endemic transmission-oriented georeferenced data attributes tend to be limited due to zero-inflated model outputs. In statistics, a zero-inflated model is a statistical model based on a zero-inflated probability distribution (i.e., a distribution that allows for frequent zero-valued arthropod-related seasonal observations) [2]. Because seasonal-sampled arthropod-related data traditionally contains excessive zero counts, these data are typically heteroscedastic. The existence of heteroscedastic residuals is a major concern in predictive linear risk modeling as it can invalidate statistical tests of significance that assume that the modeling errors are uncorrelated and normally distributed [1].

Further, to complicate matters, the number of important seasonal vector arthropod-related endemic transmission-related factors may be represented only by a few of the georeferenced explanatory observational regressors in a large empirical vector arthropod-related ecological dataset. That is, only a few log-transformed endemic transmission-oriented explanatory covariate coefficients would be able to accurately quantitate the majority of the seasonal variation in a predictive endemic

transmission-oriented risk model. For example, in an empirical seasonal-sampled dataset of vector arthropod-related larval habitat density count values, the estimators would not increase homogeneously at all sampled sites during the time-frame when the data was collected. Thus, even though seasonal aggregations of regressed habitat data attributes can reflect vector arthropod-related endemic transmission-oriented parameter estimator statistical significance levels, the input data must be properly sampled and then selected during simulation exercises to encompass any heterogeneous dependent variables. By doing so, a vector ecologist or a local abatement district manager may then extend the identification results for non-parametric simultaneous equations models as in Matzkin [3]. This would be important in situations where the observations on the vector of dependent variables might be limited, and where the number of exogeneous unobservable vector arthropod-related predictor variables is larger than the number of dependent variables.

Thereafter, new identification results for non-parametric limited dependent variable models can be robustly determined by employing the empirical sampled explanatory vector arthropod-related covariate coefficients. Commonly, these models can be subject to simultaneity in latent or observable continuous variables, and may thus depend on a large number of unobservable seasonal vector arthropod-related predictor variables. Conditions can then be provided under which the distribution of the vector of unobservable seasonal sampled vector arthropod-related endemic transmission-oriented-related predictor variables, as well as, the functions of the unobservable and observable variables are non-parametrically regressed. The residual results may thereafter be applied to a wide range of seasonal multivariate vector arthropod-related predictive risk models. In particular they can be applied to binary threshold crossing models, ordered dependent variable models and censored dependent variable models with continuous endogeneous explanatory variables with no dummy or other limited estimators. Models with “no structural shifts” considered in Heckman [4], for example, belong to this type. Examples of empirical situations that can be analyzed employing such unobservable sampled multivariate vector arthropod-related endemic transmission-oriented predictive models

include those that contain spatiotemporal variations in density count data variables (e.g., seasonal-sampled larval habitat values). Commonly, the construction of these models proceeds by first transforming the model with limited dependent variables employing a large number of unobservable variables transformed into 'observable' continuous dependent variables whereby, the same number of unobservable variables as the number of dependent variables are regressed. The transformed model may then satisfy the conditions of the residual forecasts rendered. To determine the identification of this transformed model, thereafter, the identification results for simultaneous equations models can be developed as in Matzkin [3]. As such, the identified elements in the transformed unobservable seasonal predictive multivariate vector arthropod-related endemic transmission-oriented regression-based risk model together with the particular structures of the original model can be employed for the identification of the elements of the original model. Such elements would be the particular functions and distributions generating the distribution of the unobservable vector arthropod-related seasonal-sampled predictor variables in the transformed model's structural equations. Further, the residual endemic transmission-oriented outputs would be able to describe the interaction among the latent and observable dependent variables along with the observable exogeneous variables while simultaneously quantitating the seasonal unobservable exogeneous variables.

Transforming the non-parametric model with limited dependent variables into another non-parametric seasonal predictive vector arthropod-related endemic transmission-oriented risk model with 'observable' continuous dependent variables may then extend some of the original idea as in Manski [4]. In this paper, Manski showed, among other things, that one can identify and consistently estimate the coefficients of linear subutility functions in discrete choice models without specifying parametrically the distribution of the unobservable random subutilities. Two of the key elements in Manski's [4] proof of identification were a scale normalization on the vector of coefficients and assumptions on a regression. This proof causes seasonal multivariate vector arthropod-related data to possess a strictly increasing distribution

conditional on the other regressors and a nonzero coefficient. These same assumptions may be then employed in a large number of distribution-free methods for constructing binary and multinomial robust seasonal vector arthropod-related endemic transmission-related predictive seasonal risk models. In particular, Cosslett [5] showed how these same elements may deliver as well as identify the distribution of the unobservable random term independent of the regressors. Matzkin [3] then used an additive regressor for strictly increasing conditional distribution and nonzero coefficients. Instead of the scale normalization, Matzkin specified the coefficient of this regressor to have the value one. By doing so, he was able to reveal that when the latent variable in binary threshold crossing or binary choice model is employed, the sum of this regressor plus a non-parametric function of the other regressors, plus the unobservable random, term would enable the non-parametric function and the distribution of the unobservable random, term to be identifiable. Matzkin [3] developed the results assuming that the additive unobservable random term was distributed independently of the vector of all observable explanatory variables. Lewbel [6] then showed that the continuous large support regressor need only be independent of the unobservable variables, conditional on the other explanatory dependent covariate coefficients. By so doing, this model specification would allow any seasonal-sampled vector arthropod-related endogeneous regressor to be in the identification and quantitation of a variety of unobservable endemic transmission-oriented-related data feature attributes (e.g., georeferenced district location of an aggregation of prolific larval habitats). Further, a vector ecologist or a local abatement district manager may then seasonally quantitate statistical independence between the vector of all exogeneously determined endemic transmission-oriented covariate coefficients and a vector of unobservable data attributes in the model residually forecasted estimators.

Thereafter, the connection between quasi-likelihood functions, exponential family models and the non-linear weighted least squares can be invasively examined in a seasonal predictive vector arthropod-related endemic transmission oriented risk model. Consistency and asymptotic normality of the seasonal-sampled arthropod-related data can

additionally be discussed for employing second moment assumptions. Generally, to define likelihood in such a risk model the form of distribution of the observations must be quantitated, but to define a quasi-likelihood function only a relation between the mean and variance of the endemic transmission oriented observations and the quasi-likelihood need to be employed for parsimonious uncertainty estimation. Unfortunately, for a one-parameter exponential family, the log-likelihood is the same as the quasi-likelihood [1], thus it follows that assuming a one-parameter exponential family would be the weakest sort of distributional assumption that could be efficiently constructed for a seasonally robust predictive vector arthropod-related endemic trasnmission-oriented risk model.

Fortunately, the Gauss-Newton method for calculating nonlinear least squares estimates generalizes easily to deal with maximum quasi-likelihood estimates. A rearrangement of this algorithm would produce a generalization of the uncertainty method as described by Nelder and Wedderburn [8]. The uncertainty parameter estimators may then satisfy a property of asymptotic optimality in a seasonal predictive multivariate vector arthropod-related endemic transmission-oriented risk model thus, rendering optimal properties of Gauss-Markov estimators. In statistics, the Gauss-Markov theorem states that in a linear regression model in which the errors have expectation zero and are uncorrelated and have equal variances, the best linear unbiased estimator (BLUE) of the coefficients is given by the OLS estimator [9]. Thus, suppose $Y_i = \sum_{j=1}^{K} \beta_j X_{ij} + \varepsilon_i$, for $i = 1, n, j$ are non-random but unobservable seasonal-sampled multivariate vector arthropod-related parameter estimators. The X_{ij} would then be the non-random and observable explanatory predictor variables, where ε_i and Y are random (i.e., noise). By entering this model data into SAS, a vector ecologist or an abatement district manager can include a constant in the predictive multivariate arthropod-related endemic transmission-oriented model, and then, if so desired, choose to include the variable X_K for all observed seasonal sampled values n and the residual forecasts targeting the endemic

transmission-oriented explanatory covariate coefficients, where $X_{iK} = 1$ occurs for all the quantitated regressors. The Gauss-Markov assumptions would then be $V(\varepsilon_i) = \sigma^2 < \infty$. Thereafter, employing $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$ for $i \neq j$ would quantitate any of the noise terms drawn from an uncorrelated distribution rendered from the residual forecasts targeting the statistically significant endemic transmission oriented explanatory covariates. A linear estimator of β_j is a linear combination $\hat{\beta}_j = c_{1j}Y_1 + \cdots + c_{nj}Y_n$, in which the coefficients c_{ij} are not allowed to depend on the underlying coefficients β_j, since those are not observable, but are allowed to depend on the values X_{ij}, as these data are observable [9]. The dependence of the coefficients on each X_{ij} would then be nonlinear. Commonly, the estimator is linear in each Y_i and hence, in each random ε_i, which is why this type of seasonal multivariate vector arthropod-related endemic transmission-oriented risk-based parameter estimator would be in the residual forecasts unbiasedly but if and only if, $E(\hat{\beta}_j) = \beta_j$ must represent the values of X_{ij}.

Further, by letting $\sum_{j=1}^{K} \lambda_j \beta_j$ be some linear combination of the coefficients, the mean squared error (MSE) of the corresponding predictive seasonal multivariate vector arthropod-related endemic transmission-oriented regression estimation [e.g., $E\left(\left(\sum_{j=1}^{K} \lambda_j(\hat{\beta}_j - \beta_j)\right)^2\right)$] would be the expectation of the square of the weighted sum across the sampled parameter estimators based on the differences between the estimators and the corresponding wxplanatory covariate coefficients to be regressed. Since commonly a vector ecologist and/or a local abatement district manager would assume that all the parameter estimates are unbiased, this MSE would be the same as the variance of the linear combination. The BLUE of the vector β of parameters β_j would then be associated with the smallest MSE for every vector λ based on the linear

combination parameters in the predictive model. This would then be equivalent to the condition that $V(\tilde{\beta}) - V(\hat{\beta})$ is a positive semi-definite matrix for every other linear unbiased estimator $\tilde{\beta}$ in the residual forecasts as well.

Interestingly, a predictive seasonal multivariate vector arthropod-related endemic transmission-oriented risk model would be considered semipositive-definite (or sometimes nonnegative-definite), if $x^* Mx \geq 0$ for all x in C^n or, all x in R^n is employed for constructing the uncertainty matrix. A matrix M is positive-semidefinite, if and only if, it arises as the Gram matrix of some set of vectors [1]. In contrast to the positive-definite case, these vectors need not be linearly independent. For any matrix A, the matrix A^*A is positive semidefinite, and $\text{rank}(A) = \text{rank}(A^*A)$ [1]. Conversely, any seasonal multivariate vector arthropod-related Hermitian positive semidefinite matrix M can be written as $M = A^*A$ (i.e., Cholesky decomposition). In linear algebra, the Cholesky decomposition or Cholesky triangle is a decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose [8]. However, a predictive seasonal multivariate vector arthropod-related risk model Hermitian matrix is positive semidefinite, if and only if, all of its principal minors are nonnegative. It is thus not enough to consider the leading principal minors only as is checked on the diagonal matrix with entries 0 and −1. Conversely, the predictive seasonal multivariate vector arthropod-related endemic transmission oriented risk model would be considered negative-semidefinite, if, $x^* Mx \leq 0$ for all x in C^n or, all x in R^n the uncertainty matrix. The OLS regression would then be based on the function $\hat{\beta} = (X'X)^{-1}X'Y$ of Y and X, where X' denotes the transpose of X which then subsequently would minimize the sum of squares of residuals (e.g., predictions) using: $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{K}\hat{\beta}_j X_{ij}\right)^2$. The main idea of employing the OLS in a multivariate seasonal vector arthropod-related predictive model is that the least-squares estimator would be

uncorrelated with every linear unbiased estimator of zero (i.e., with every linear combination $a_1Y_1 + \cdots + a_nY_n$, whose coefficients do not depend upon the unobservable β but whose expected value is always zero) [1].

Thus, vector ecologists and local abatement district managers may employ a vast number of multivariate probability distributional approaches for quantitating seasonal-sampled vector arthropod-related endemic transmission-oriented explanatory covariate coefficients. Unfortunately, regardless of which statistical package is employed for constructing a seasonal predictive multivariate endemic transmission risk model, generally, there would be too much redundant information in the empirical datasets of the vector arthropod-related georeferenced spatial data feature attributes to construct a robust covariance matrix. Commonly a covariance matrix in a predictive vector arthropod-related model is represented as square matrix whose diagonal entries are the variances and whose off diagonal entries are the covariances between, the row/column labeling the endemic transmission oriented variables. For example, pseudospatially replicated data commonly arises from regressed seasonal-sampled larval density count values as the sampled productive habitats commonly share similar predictors due to spatial "nearness" [2]. As such, the multiple regression assumptions in the seasonal predictive arthropod-related endemic transmission-oriented uncertainty risk model matrix would be violated. The variation in a collection of vector arthropod-related random points in two-dimensional space thus may not be characterized fully by a single number, nor would the variances in the x and y directions contain all of the necessary information for efficient quantification of the sampled explanatory covariate coefficients. These departures from normality would include that the seasonal sampled arthropod-related endemic transmission-oriented risk model residual forecast errors are normally distributed and that these errors have a constant predictive variance.

Fortunately, redundant information in a seasonal predictive vector arthropod-related endemic transmission-oriented risk model can be seasonally quantitated based on projections in geospace of probability distributions. Spatial autocorrelation measures the correlation of a variable with itself through geospace [9]. The correlation among seasonal

sampled multivariate vector arthropod-related explanatory covariate coefficient values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface will introduce a deviation from the independent observations assumption of classical statistics. Quantitating regressional replicated seasonal-sampled multivariate vector arthropod-related endemic transmission-oriented explanatory covariate coefficients in geospace using autocorrelation statistics may enable accurate formalized non-negative decomposition of the duplicated data into standard errors of prediction.

Currently, the standard method for analyzing regional ecological spatial autocorrelation trends in seasonal-sampled vector arthropod-related multivariate empirical datasets is the Mantel test. This test can spatially quantitate the overall relationship between distance and similarity between sampled sites (e.g., georeferenced vector arthropod-related larval habitats). Typically, when performing a Mantel test, a matrix consisting of the measurements between all pairs of sampled sites (e.g., larval habitats) and a matrix consisting of the similarity between the sampled values across all pairs of the seasonal-sampled arthropod-related georeferenced sites are employed for determining "hot spots". Thereafter, by plotting the correlation coefficients against a known vector arthropod-related Euclidean distance measurement (e.g., distance between a sampled georeferenced larval habitats and the epidemiological capture point), a robust empirical predictive model could be constructed based on randomization or permutation tests for determining endemic zones of importance (e.g., hyperendemic) and their seasonal-sampled transmission-oriented explanatory covariate coefficients. For example, the correlation between two empirical vector arthropod-related larval habitat datasets could be calculated and the measure of correlation reported using the test statistic on which the seasonal predictive multivariate endemic transmission-oriented risk model would be based upon.

In principle, any correlation coefficient could be employed for constructing a robust seasonal multivariate predictive vector arthropod-related endemic transmission-oriented risk model but normally the Pearson product-moment correlation coefficient is most commonly used. A robustly quantized Pearson's correlation coefficient generated between

any two georeferenced vector seasonal-sampled arthropod-related endemic transmission-oriented observational predictors could be then defined as the covariance of the two estimators divided by the product of their standard deviations. The formula for ρ in the seasonal predictive multivariate vector arthropod-related model would then be $\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$ [2]. Thereafter, by obtaining a formula for r and by substituting the seasonal vector arthropod-related predictive risk estimates of the covariances and variances based on a sample into the Pearson's correlation coefficient formula [i.e., $r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$], an equivalent expression could be rendered for the predictive squared correlation coefficient. Further, based on a sample of paired seasonal vector arthropod-related larval habitat spatial data feature attributes (X_i, Y_i), for example, the sample Pearson correlation coefficient could be written as $r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{(X_i - \overline{X})}{s_X} \right) \left(\frac{(Y_i - \overline{Y})}{s_Y} \right)$, where $\frac{X_i - \overline{X}}{s_X}$, $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, and $s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$ for synthetizing unbiased optimal residual forecasts from the regressed endemic transmission-oriented seasonal-sampled explanatory covariates.

Presently, many statistical packages include routines for carrying out the Mantel test. For example, employing **ade4** library, a vector ecologist or a local abatement district manager can perform a robust Mantel test in R, a free software environment for computing predictive models and graphics (http://www.rproject.org/). Unfortunately, to run a Mantel test accurately for quantitating latent autocorrelation uncertainty coefficients in residual forecasts for optimally targeting seasonal vector arthropod-related endemic transmission-oriented explanatory covariate correlation coefficients, there would be a necessity to generate two distance-based matrices: one containing spatial distances and the other one containing distances between measured outcomes at a sample point (e.g.,

georeferenced capture point). In the measured outcome matrix then entries for pairs of sampled georeferenced vector arthropod-related larval habitat locations with similar outcomes (e.g., larval density counts) may be then determined to be lower than for pairs of sampled georeferenced larval habitats with dissimilar values. This can be further tested employing the **dist** function in R also. Routinely, the Mantel model function would perform two distance correlation tests employing the **mantel.r test**. The first test would consist of calculating the correlation of the entries in the matrices and then permuting the matrices by calculating the same test statistic under each permutation. The second test would then compare the original test statistic to the distribution of test statistics in a seasonal predictive multivariate vector arthropod-related risk endemic transmission-oriented model from the permutations to generate a *p*-value. The number of permutations would then define the precision with which the risk model residually forecasts *p*-values.

The number of permutations in a robust multivariate seasonal vector arthropod-related endemic transmission-oriented predictive risk model can then be specified by a vector ecologist or a local abatement manager by employing a default of 99, for example. By so doing, a collection of n distinguishable vector arthropod-related seasonal-sampled objects (e.g., larval habitats) values can be provided by quantiting a permutation relationship: $n^{P}r = \frac{n!}{(n-r)!}$ [1]. Thereafter, based on the residual forecasts, the null hypothesis in the risk model may be accepted or rejected. Validation exercises can then be based on the covariance matrices, spatial distance, and sampled larval habitats distance measurements that may or may not be unrelated when alpha = 0.05, for example. The observed correlation in the seasonal predictive vector arthropod-related endemic transmission-oriented risk model pseudo R^2 may then suggest that the matrix entries are positively/negatively associated in geospace. Further, smaller differences in the predictive seasonal risk model estimators may be observed among sampled pairs of stations (e.g., georeferenced larval habitats) that are close to each other as compared with those habitats that are farther from each other.

Note that since this test would be based on random permutations, the same code will always arrive at the same observed correlation but, fortunately rarely of the same p-value in an ecological dataset of seasonal predictive vector arthropod-related endemic transmission oriented regression-based risk validation model residual forecasts. Instead, the computation would yield a Z statistic: $Z = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij}$, where A and B would represent the distance matrices. More commonly, this statistic would be normalized via a standardized normal transformation where the mean of the seasonal predictive vector arthropod-related regression-based uncertainty matrix is subtracted from each element. Thereafter, each element would be divided by the standard deviation. This then would yield an r statistic: [i.e., $r = \frac{\sum\sum stdA_{ij}stdB_{ij}}{n-1}$]. The significance of the test statistic derived from the seasonal predictive vector arthropod-related risk based endemic transmission-oriented model could then be assessed by one of two methods. Commonly, a permutation test is performed if the empirical datasets is small in sample size $(n < 20)$. Conversely, for large sample sizes (e.g., a seasonal predictive vector arthropod-related endemic transmission-oriented risk model with an $n > 100$) significance of the estimators may be determined by an asymptotic t-approximation where the test statistic is transformed into a t statistic. A significant result may then indicate spatial autocorrelation in the residual forecasts targeting the endemic transmission-oriented explanatory covariates.

Unfortunately, Mantel tests have two disadvantages for performing a robust space-time seasonal multivariate predictive vector arthropod-related endemic transmission-oriented autocorrelation-based risk analyses. First, the selection of critical space-time distances for data transformation for the Mantel test is subjective. Second, since the Mantel statistic is the sum of the products of space and time distances, only sampled coefficients linear in form should be expected in the contagious processes of a robust predictive vector arthropod-related risk model. Further, the test would not be sensitive to non-linear associations

between small space and time distances in the risk model residual forecasts. Additionally, Mantel tests have poor performance compared to alternative methods, including those that employ low power. As such, under inflated type-I error features frequently occur in vector arthropod-related predictive risk models. A remedy for the inflated type-I error of three-way Mantel tests for quantitating the sampled vector arthropod-related endemic transmission-oriented explanatory covariate coefficients may be performed by using permutations, however, this test may display considerably lower power than independent contrasts. Thus, the use of the Mantel tests for constructing a robust seasonal predictive vector arthropod-related endemic transmission-oriented risk model may render biased misspecified residual forecasts. This situation may be amplified in cases in which the seasonal-sampled data feature attributes are expressed as pairwise distances in the datasets of the seasonal-sampled vector arthropod-related data attributes (e.g., Euclidean distances between prolific sampled larval habitats) for targeting statistically significant seasonal multivariate endemic transmission-oriented explanatory covariate coefficients.

The k nearest neighbour (k-NN) statistic is another technique whereby, the number of case pairs that are k nearest neighbours in both space and time are evaluated under the null hypothesis of independent space for constructing a robust seasonal multivariate predictive vector arthropod-related endemic transmission-oriented model. The test may be applied for quantitating nearest neighbour relationships associated to georeferenced endemic transmission-oriented larval habitat explanatory covariate coefficients, for example. The k-nearest neighbour algorithm is among the simplest of all machine learning algorithms [1]. In this algorithm, the seasonal-sampled vector arthropod-related object (e.g., georeferenced larval habitat) would be assigned to the class most common among its k nearest neighbours where, k is a positive integer. Thus, if $k = 1$, in a predictive multivariate seasonal vector arthropod-related risk model is assigned to the class of its nearest neighbouring georeferenced larval habitat endemic transmission-oriented explanatory covariate, the estimators may be spatially delineated based on sampled coefficient values. The same method can then be employed for regression

of the sampled explanatory uncertainty covariate coefficients by simply assigning the property value for a specific vector arthropod-related object (e.g., larval habitat capture point) as the average of the seasonal-sampled values, for example, based on its k nearest neighbours. The algorithm may be also useful for determining significant seasonal-sampled parameter estimators since the weight of the contributions of the neighbours derived from nearer sampled larval habitats neighbors may be deemed to contribute more to disease transmission than the more distant habitat neighbors.

Importantly, employing k-NN as a common weighting scheme for constructing a robust predictive multivariate seasonal vector arthropod-related endemic transmission-oriented risk model would render each neighbour a weight of $1/d$, where d is the distance to the neighbour. This scheme would be a generalization of a weighted matrix. The neighbours (e.g., georeferenced larval habitats) can then be taken from a set of objects for which the correct classification or, in the case of regression, the value of the property, is known. This can be thought of as the training set for the algorithm for efficiently quantitating an empirical dataset of seasonal-sampled vector arthropod-related georeferenced endemic transmission-oriented explanatory covariate coefficients although no explicit training step would be actually required. The k-NN would then determine sensitivity to the space-time interaction patterns expected and exhibited by the regressed endemic transmission-oriented cluster-based covariates. Fortunately, the k-NN algorithm would not require parameters such as critical distances between the sampled predictor variables to be estimated and thus, the algorithm, may be further employed to test various hypotheses about the spatial and temporal scale of any significant clustering processes. By doing so, the k-NN method would address significant weaknesses in the existing space-time predictive multivariate seasonal vector arthropod-related risk model cluster tests as well. In pattern recognition, the k-NN is a non-parametric method for classifying objects based on closest training examples in feature space [4]. The k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification [1]. As such, the

k-nearest neighbour algorithm would be also sensitive to the local structure of the seasonal-sampled latent serially correlated vector arthropod-related spatial data feature attributes. Further, nearest neighbour rules would be able to implicitly compute the decision boundary in the endemic transmission-oriented risk model.

Unfortunately, the nearest-neighbor method for predicting vector arthropod-related data suffers severely from the "curse of dimensionality". The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low-dimensional settings, such as the three-dimensional physical space [2]. The problem of lower dimensionality in a seasonal predictive multivariate vector arthropod-related endemic transmission-oriented risk model is that when the dimensionality increases, the volume of the space in the model would increase so fast that the available data becomes sparse. This scarcity would be problematic for seasonally delineating statistically significant endemic transmission-oriented explanatory covariates accurately since this would require determining significance of the each model parameter estimator. Unfortunately, quantitating predictive seasonal-sampled multivariate arthropod-related data often relies on detecting areas where larval habitats form groups with similar properties (e.g., prolific larval density counts) in high dimensional geospace. Subsequently, this would then require a robust weighted matrix for accurate spatial regressive quantitation and precise parameter significance estimation. Unfortunately, due to excessive latent autocorrelation in the forecasts in the dataset of non-linear empirical sampled datasets, the georeferenced explanatory uncertainty covariate coefficients would be misspecified. As such, the residual forecasts can appear to be sparse and dissimilar in many ways preventing common data organization strategies from being efficient for targeting important endemic transmission-oriented seasonal-sampled observational explanatory predictors.

Another effect of high dimensionality on distance functions concerns *k*-nearest neighbour (*k*-NN) graphs constructed from a regressed dataset of predictive seasonal-sampled multivariate arthropod-related empirical-

sampled explanatory covariates for precisely quantitating distance functions. For example, as dimensionality increases in degree distribution, the k-NN seasonal predictive multivariate endemic transmission-oriented risk model residually forecasted derivatives will skew to the right. In probability theory and statistics, skewness is a measure of the extent to which a probability distribution of a real-valued random variable "leans" to one side of the mean which can be positive or negative, or even undefined [1]. Statistical analyses of the output from a nonlinear seasonal predictive vector arthropod-related model must then take into consideration that such model parameters as (i.e., mean, variance, autocorrelation, and skewness) can be interdependent quantities, particularly as the system becomes less stable [2].

Griffith [9] implements spatial autoregressive models employing SAS's nonlinear procedure PROC NLIN. This estimation is nonlinear because: (i) the Jacobian term, is a function of the spatial autocorrelation parameter which would then appear as a divisor of each predictive regression risk model term; and (ii) each coefficient appearing in the product term is multiplied by the spatial autocorrelation parameter. In vector calculus, the Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function [2]. For example, suppose $F : \mathbb{R}^n \to \mathbb{R}^m$ is a function which takes as input real n-tuples in a seasonal multivariate predictive vector arthropod-related endemic transmission-oriented risk model which produces real m-tuples in the residual forecasts. A function can then be provided by m a seasonal-sampled dataset of larval habitat component functions, $F_1(x_1, \ldots, x_n)$, $\ldots$, $F_m(x_1, \ldots, x_n)$. Thereafter, partial derivatives of all these functions with respect to the seasonal arthropod-related endemic transmission-oriented explanatory predictor variables $x_1, \ldots, x_n$, can be organized in an m-by-n matrix. The Jacobian matrix J of F would then be illustrated as $J = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}$. This matrix can thereafter reveal functions

of $x_1, \ldots, x_n$, whose entries can be part of a robust seasonal multivariate arthropod-related endemic transmission-oriented predictive regression-based equation. The residual forecasts targeting the seasonal endemic transmission-oriented explanatory covariates would then be denoted by $J_F(x_1, \ldots, x_n)$ and $\frac{\partial(F_1, \ldots, F_m)}{\partial(x_1, \ldots, x_n)}$. Further, according to the inverse function theorem, the matrix inverse of the Jacobian matrix of an invertible function in a predictive seasonal vector arthropod-related model would be the Jacobian matrix of the inverse function.

In mathematics, specifically differential calculus, the inverse function theorem gives sufficient conditions for a function to be invertible in a neighborhood of a point in its domain [1]. The theorem also provides a formula for the derivative of the inverse function. An inverse function is a function that undoes another function [1]. If an input x into the function f produces an output y, in a seasonal predictive multivariate vector arthropod-related endemic transmission model then putting y into the inverse function g produces the output x, and vice versa, i.e., $f(x) = y$ and $g(y) = x$. More directly, if $g(f(x)) = x$, then $f(x)$ leaves x unchanged in the risk model [1]. A function f in a seasonal multivariate predictive vector arthropod-related endemic transmission- oriented risk model that has an inverse function would then be invertible. The inverse function in the model would be thereafter uniquely determined by f and/or by f^{-1}. Instead of then considering the inverses for individual inputs and outputs in the predictive multivariate seasonal vector arthropod-related risk model residually forecasted derivatives, the function would instead send the whole set of inputs of the domain to a set of range outputs. If f is a function whose domain is the set X, and whose range is the set Y in the vector arthropod-related model, f would then be invertible in the residual forecasts for robustly statistically targeting the important endemic transmission-oriented explanatory covariates, but only if there exists a function g with domain Y and range X with the property: $f(x) = y \Leftrightarrow g(y) = x$. Further, since f is invertible, the function g would be unique in the predictive seasonal vector arthropod-related risk model when there is only one function g satisfying the model

outputs. That function g would then be the inverse of f and would then be denoted as f^{-1} in the residual forecasts. Stated otherwise, a function is invertible in a predictive multivariate seasonal vector arthropod-related endemic transmission- oriented risk model, if and only if, its inverse relation is a function on the range Y, in which case the inverse relation would be the inverse function.

It is important to remember that not all functions have an inverse. For this rule to be applicable thus, in a robust seasonal predictive multivariate vector arthropod-related endemic transmission-oriented risk model, each element $y \in Y$ must then correspond to no more than one $x \in X$. A function f with this property is called one-to-one, or information-preserving, or an injective function [1]. In actuality, a function that preserves distinctness can never map distinct elements of its domain to the same element of its codomain. In other words, every element of the function's codomain would be the image of at most one element of its domain in the risk model residual forecasts. If in addition, all of the elements in the codomain are in fact the image of some element of the domain in a seasonal predictive robust multivariate vector arthropod-related endemic transmission oriented model, then the function would be a bijective function.

Bijection (or bijective function or one-to-one correspondence) is a function giving an exact pairing of the elements of two sets [1]. As such, in a predictive seasonal vector arthropod-related risk model, every element of a seasonal-sampled dataset would be paired with exactly one element of another seasonal-sampled dataset and every element of the other dataset would be paired with exactly one element of the first sampled dataset. There can be no unpaired elements for proper spatiotemporal quantitation of georeferenced data variables [1]. In formal mathematical terms, a bijective function $f : X \to Y$ then would be one-to-one mapping of a set X to a set Y. A bijection from the set X to the set Y has an inverse function from Y to X [1]. If X and Y are then finite sets in a seasonal predictive vector arthropod-related endemic transmission oriented risk model, then by bijection, they would have the same number of elements. For extremely large seasonal-sampled dataset ($n > 100$), the

picture unfortunately is more complicated, leading to the concept of including a cardinal number, which is actually a way to distinguish various sizes of infinite sets. A bijective function from a set to itself is also called a *permutation* [1]. Bijective functions are essential to many areas of mathematics including the definitions of isomorphism, homeomorphism, diffeomorphism, permutation group, and projective mapping (e.g., predictive seasonal vector arthropod-related endemic transmission-oriented risk modeling).

As such, if the Jacobian of the function $F : \mathbf{R}^n \to \mathbf{R}^n$ in a seasonal predictive multivariate vector arthropod-related endemic transmission-oriented risk model is continuous and nonsingular at the point p in $\mathbf{R}^n$, then F would be invertible when restricted to some neighbourhood of p and $(J_F^{-1})(F(p)) = [(J_F)(p)]^{-1}$. The spatial structuring would thereafter be achieved by constructing a linear combination of a subset of the eigenvectors of a modified geographic weights matrix, using $(\mathbf{I} - \mathbf{11}'/n)\mathbf{C}$ $(\mathbf{I} - \mathbf{11}'/n)$ that appears in the numerator of the Moran's coefficient (MC).

Moran's coefficient (MC) tests can quantitate global spatial autocorrelation coefficients in continuous data [1]. These spatial statistics can be based on cross-products of the deviations from the mean and can be calculated for n sampled predictive multivariate vector arthropod-related seasonal observations, for example, on a variable x at a sampled vector georeferenced larval habitat locations i, j as: $I = \dfrac{n}{S_0} \dfrac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$, where $\bar{x}$ is the mean of the x variable, w_{ij} are the elements of the weighted matrix, and S_0 is the sum of the elements of the weight matrix: $S_0 = \sum_i \sum_j w_{ij}$. Moran's I is similar but not equivalent to a correlation coefficient. It varies from –1 to +1. In the absence of autocorrelation and regardless of the specified weight matrix, the expectation of Moran's I statistic would be $-1/(n-1)$, in a seasonal

predictive vector multivariate arthropod-related endemic transmission-oriented-related risk model which would tend to zero as the sample size in the empirical sampled ecological dataset increases. For a row-standardized spatial weight matrix, the normalizing factor S_0 equals n (i.e., since each row sums to 1), and the statistic simplifies to a ratio of a spatial cross-product to a variance [2]. A Moran's I coefficient larger than $-1/(n-1)$ in the predictive multivariate seasonal residual forecasts targeting endemic transmission zones would then indicate positive spatial autocorrelation (PSA) [i.e., spatial aggregation] of a cluster based on similar sampled values (e.g., larval density counts); Conversely, Moran's I less than $-1/(n-1)$, would then indicate negative spatial autocorrelation (NSA) (i.e., dissimilar larval habitat values clustering in geospace).

Thereafter, significance levels can be obtained by studying the matrix form of the MC in a seasonal multivariate predictive endemic transmission-oriented risk model specifically employing the term $\mathbf{Y}T(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)\mathbf{Y}$ corresponding to the first summation the algorithm where $\mathbf{I}$ is an n-by-n identity matrix, $\mathbf{1}$ is an n-by-1 vector of ones, T is the matrix transpose operation, and $(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)$ is the projection matrix that centers the vector $\mathbf{Y}$ [10]. The extreme eigenvalues of matrix expression $(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)$ may then determine the range of the modified correlation coefficient in the ecological dataset of seasonal predictive multivariate arthropod-related endemic transmission-oriented risk model residual forecasts. The extreme eigenvalues of matrix expression $(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}T/n)$ could then also determine the range of the modified correlation coefficient by using MC to quantitate latent NSA and PSA.

Geary's C statistic is based on the deviations in responses of each observation with one another: $C=\dfrac{n-1}{2S_0}\dfrac{\sum_i\sum_j w_{ij}(x_i-x_j)^2}{\sum_i (x_i-\bar{x})^2}$, which ranges from 0 (i.e., maximal positive autocorrelation) to a positive value

for high NSA coefficients [11]. Its expectation would be 1 in a robust seasonal predictive multivariate vector arthropod-related endemic transmission-oriented model in the absence of autocorrelation regardless of the specified weight matrix. If the value of Geary's C is less than 1 in the endemic transmission-oriented risk model, for example, it would indicate PSA. Moran's I is a more global measurement and sensitive to extreme values of x, whereas, Geary's C is more sensitive to differences in small neighbourhoods [2]. In general, Moran's I and Geary's C result in similar conclusions. However, Moran's I would be preferred in predictive seasonal multivariate vector arthropod-related risk mapping since Griffith [9] revealed that Moran's I is consistently more powerful than Geary's C. Regardless, there is no straightforward way to calculate Moran's I and Geary's C for spatially targeting endemic transmission-oriented covariate coefficients by only using SAS for statistically targeting significant explanatory covariate covariates.

Fortunately, freely available SAS code for the MC models, spatial random effects models, cluster detection, spatial diffusion, and much more can be found on-line (www.sas.edu). One source of code using PROC IML for robust autoregressive seasonal multivariate predictive vector arthropod-related endemic transmission-oriented risk mapping can be found at the website of E. B. Moser at the Department of Experimental Statistics, Louisiana State University (http://www.stat.lsu.edu/faculty/moser/spatial/spatial.html). Unfortunately, the p-values calculated by "TESTOFI" and "TESTOFC" in the residual forecasts from a seasonal multivariate predictive vector arthropod-related endemic transmission-oriented risk model may be one-sided in PROC NLIN. As such, there may be a cause for concern in "TESTOFC", for proper quantitation of seasonal sampled vector arthropod-related data as it is calculated by using the same variance for two tests employing normality and randomization assumptions. Although the operating environment in which this particular version of SAS runs and the precision of the algorithms will not change, problems with the PROC NLIN. Model convergence can render biased error estimates.

Recently, SAS has included many specific geographical functions for optimal predictive multivariate seasonal risk mapping vector arthropod-related endemic transmission oriented georeferenced explanatory

covariate coefficients. The SAS/GIS and SAS/GRAPH software provide many mapping capabilities within SAS (see SAS/GIS 2008 and SAS/GRAPH 2008 for details about the software and its procedures). Also, SAS has implemented geostatistical procedures like PROC VARIOGRAM, which includes an option for computing Moran's I and Geary's C statistics using binary, row standardized or distance weighted matrices. Also available is spatially structured random effects intercept/option models based on a geostatistical semivariograms, using statements like *repeated/sub = intercept type = SP* (*EXP*) (*U V*), where *EXP* is the exponential characterization of semivariance and (*U V*) are geographic coordinate pairs. A spatially structured random effects intercept also can be specified for a predictive seasonal multivariate vector arthropod-related risk model without this built-in geostatistical option. As such, the residual forecasts targeting specific endemic transmission-oriented explanatory covariate coefficients and the other vital estimators including selected eigenvectors may be revealed in the model statement by specifying *random intercept/type = VC sub = ID*.

Probably the most powerful use of SAS in spatial statistics for seasonal predictive multivariate vector arthropod-related endemic transmission-oriented risk mapping however, would be the ability to modify existing procedures to include spatial information in the residual forecasts. For example, PROC NLIN can be employed with weights to estimate any valid seasonal vector arthropod-related semivariogram by using the output of PROC VARIOGRAM. A spatial regression can then be specified with a variety of different methods in PROC NLIN. Further, PROC GENMOD can be used for generating generalized linear model (GLM) regression specifications from an ecological dataset of predictive robust endemic transmission-oriented seasonal multivariate risk model parameters by including eigenvector spatial filter proxy variables in the regression. Wang et al. [10] gives an example of wasteful commuting and sample SAS code using PROC LP to solve linear programming problems. These procedures, along with the flexibility of PROC IML, SAS's interactive matrix language and the data step may enable the customized programming of many standardized quantitative geographical predictive seasonal multivariate vector arthropod-related predictive risk models, such as the Huff model, the Grain-Lowry model, and the doubly constrained gravity model.

Currently, the % GLIMMIX macro, available in the SAS/STAT® sample library, may extend the mixed model technology of PROC NL MIXED to a more robust generalized linear mixed model (GLMM). Originally available as a Webdownload for Windows and several UNIX platforms for SAS 9.1®, PROC GLIMMIX has been updated for SAS 9.2 (www.sas.edu). The GLIMMIX procedure is an add-on for the SAS/STAT® in SAS® on the Windows platform, which is currently downloadable (support.sas.com). The GLIMMIX install creates an SAS® program called GLIMMIX_TPL.SAS. Meantime, the % NLINMIX macro, also available in the SAS/STAT® sample library may provide a similar framework for non-linear mixed seasonal multivariate vector arthropod-related endemic transmission-oriented predictive risk model construction. The GLIMMIX procedure fits statistical models to data with correlations or noncontact variability especially where the response is not necessarily normally distributed. These generalized predictive models like linear mixed models would initially assume normal (i.e., Gaussian) random effects in the residual forecasts targeting the statistically significant vector arthropod-related endemic transmission-oriented explanatory covariate coefficients. Conditional on these random effects, the seasonal-sampled data may have any distribution in the exponential family. This program can also create the templates needed to format the GLIMMIX procedure for constructing a robust vector arthropod-related endemic transmission-oriented predictive uncertainty model. By doing so, SAS models can extend a mixed seasonal-sampled predictive endemic transmission-oriented distribution model for more accurately delineating and seasonally targeting statistically significant explanatory covariate coefficients.

For example, the ability to include data from non-Gaussian distributions rendered after regressing seasonal sampled endemic transmission-oriented explanatory covariates can implement low-rank smoothing based on SAS programming. Kalman filltering-smoothing is a fundamental tool in statistical time series vector arthropod-related predictive risk modeling. Standard implementations of the Kalman filter-smoother would require $O(d3)$ time and $O(d2)$ space per time step, where d is the dimension of the state variable in a high-dimensional predictive multivariate seasonal vector arthropod-related endemic transmission

oriented risk model Although the estimators would be approximated in terms of a low-rank perturbation based on a prior state covariance matrix in the absence of any observation the filter would be able to compute distributional effects, or to define link and variance functions in a seasonal multivariate predictive vector arthropod-related endemic transmission-oriented risk model. By so doing, the model may render robust residual forecasts of endemic transmission-oriented explanatory covariate coefficients in an epidemiological interventional study site.

Low rank radial smoothing using GLIMMIX is a semiparametric approach to smooth curves [SAS Institute, statistical analysis with the GLIMMIX procedure course notes, SAS Press]. Low rank radial smoothing using GLIMMIX is actually a semiparametric approach to smooth curves. Specifying TYPE=RSMOOTH option in RANDOM statement, can help implement a spline smooth approach, for example in a robust seasonal predictive arthropod-related endemic transmission-oriented risk model. In such a framework, data preparation could be performed extremely easily by employing the OUTDESIGN= & NOFIT options in v9.2 PROC GLIMMIX. Thereafter, by employing PROC SCORE twice on the design matrix, a vector ecologist or a local abatement district manager can score the fixed effects design matrix X and the random effects design matrix Z, respectively, by adding the score from this radial smoothing method in any seasonal multivariate vector arthropod-related endemic transmission oriented predictive risk model. This can be formulated in SAS as:

```
Proc glimmix data=train_data absconv=0.005;
    Model y = &covars /s;
    Random &z /s type=rsmooth knotmethod=equal (20);
Run;

Proc glimmix data=test nofit outdesign=test2;
    Model y = &covars /s;
    Random &z /s type=rsmooth knotmethod=equal (20);
Run;
```

```
Pros score data=test2 score=beta fix type=prams out=score fix;
    Vary &covers;
Run;

Proc score data=test2 score=beta_random type=prams out=score_random;
    Var _z:
Run;
```

By doing so, the model fit by the GLIMMIX procedure would then extend the GLM-related seasonal multivariate vector arthropod-related endemic transmission-oriented predictive model by incorporating correlations among the responses. This can be accomplished easily by including random effects in the linear predictor and/or by modeling the correlations among the sampled endemic transmission-oriented observational explanatory covariate coefficients directly. Fortunately, the GLIMMIX procedure distinguishes two approaches; the "G-side" and "R-side" random effects (www.sas.edu). This terminology draws on common specification of a linear mixed model, for example, where the random effects have a normal distribution with mean 0 and variance matrix *G*. The distribution of the errors then in the predictive multivariate seasonal arthropod-related endemic transmission-oriented risk model would be normal with mean 0 and variance R. Modeling with G-side effects thereafter will then specify the columns of the *Z* matrix and the structure of *G* in the residual forecasts targeting the statistically significant vector arthropod-related endemic transmission-oriented explanatory covariate coefficients. Thus, predictive risk modeling with R-side effects can directly specify the covariance structure of the *R* matrix in a predictive seasonal multivariate vector arthropod-related risk-based seasonal analyses.

In an SAS-derived generalized linear mixed seasonal multivariate vector arthropod-related endemic transmission-oriented predictive risk model, G-side random effects could also be constructed by adding random effects to any predictor. This would then lead to a model of the form $g(E[Y]) = x0\ \_+ z0,$ where again the normal distribution would have a

mean 0 and variance matrix, which would be *G*. Instead of specifying a distribution for *Y* in the seasonal multivariate predictive vector arthropod-related endemic transmission-oriented risk model, as in the case of a GLM, a distribution for the conditional response, *Y* may be specified instead as a conditional model specification. A model with only R-side random effects is known as a marginal model because, there are no random effects on which you can condition the response [1]. Thus, in a marginal seasonal-sampled multivariate vector arthropod-related predictive regression-based model, the mean $g(E[Y]) = g(\mu) = x0$ may be specified in the residual forecasts targeting the endemic transmission-oriented explanatory covariate coefficients by quantitating the covariances among the *Yi*. The distributional assumption would remain relevant as the mean and the variance in the residual forecasts would be functionally related to most distributions in the exponential family. For example, a Poisson-related predictive seasonal multivariate vector arthropod-related regression-based risk model distribution $[\text{i.e., } \text{var}[Y] = E[Y] = \mu]$, can be determined if, *A* is a diagonal matrix in the residual forecasts targeting the endemic transmission-oriented explanatory covariate coefficients containing the variance functions encompassed in the regression. As such, the variance matrix in the predictive risk model with only R-side random components would be $\text{var}[Y] = A1/2RA1/2$.

Thereafter, by combining G-side and R-side random effects in the seasonal multivariate predictive arthropod-related risk model, statistically significant seasonal endemic transmission-oriented explanatory covariate coefficients may be further quantitated. For example, in an epidemiological interventional study site in which randomly sampled vector arthropod-related larval habitats are measured repeatedly over time, a vector ecologist and/or abatement district manager may regress sampled larval habitat effects (e.g., seasonal density count) by including random intercepts. These would be the G-side components as they would contribute to any sampled observational predictor. For a given sampled vector arthropod-related endemic transmission-oriented spatial feature attribute (e.g., larval habitat density count), then the correlations over time can be modeled with an R-side

autoregressive structure. Combining these elements, the predictive risk model fit by the GLIMMIX procedure can then be represented as follows: $E[Y|] = g - 1(X\_+ Z) = g - 1(\_) = \mu \text{ var}[] = G \text{ var}[Y|] = A1 / 2RA1 / 2,$ where $g - 1(\cdot)$ is the inverse link function. By so doing, the class of generalized linear predictive seasonal arthropod-related multivariate mixed models would contain several other important types of statistical information (e.g., random effects, identity link functions, and normality of distribution) in the residual forecasts for precisely targeting the statistically important endemic transmission-oriented covariates.

Additionally, the GLIMMIX procedure could provide Laplace and adaptive quadrature estimation methods with which a likelihood-based empirical estimator in a robust predictive seasonal vector arthropod-related endemic transmission-oriented model could be quantitated. Aspects common to adaptive quadrature and Laplace approximation for the seasonal vector arthropod-related predictive regression-based framework could then include estimated precision rates, which then could be denoted in a second derivative matrix $H = -\frac{\partial^2 \log\{L(\beta, \theta, \hat{\gamma})\}}{\partial[\beta, \theta]\partial[\beta', \theta']}$ and thereafter evaluated at the converged solution of the optimization process. Partitioning its inverse as $H^{-1} = \begin{bmatrix} C(\beta, \beta) & C(\beta, \theta) \\ C(\theta, \beta) & C(\theta, \theta) \end{bmatrix}$, the METHOD=LAPLACE and METHOD=QUAD would then render seasonal SAS-derived multivariate predictive vector arthropod-related seasonal risk model spatial feature attributes for further mathematically targeting endemic transmission-oriented explanatory covariate coefficients. The GLIMMIX procedure would compute H by finite forward differences based on the analytic gradient of $\log\{L(\beta, \theta, \hat{\gamma})\}$. The partition $C(\theta, \beta)$ would then serve as the asymptotic covariance matrix of the covariance parameter estimates employing an ASYCOV option in the PROC GLIMMIX statement. The asymptotic covariance matrix is the covariance matrix of parameter estimates which is also known variously as the inverse of the Fisher information matrix (denoted $I(q)^{-1}$). Elements along the diagonal can then be used to represent the variance expected of

each asymptotic covariance matrix parameter estimate over repeated sampling, and can be interpreted as indices of precision of estimation. Off-diagonal elements represent covariances of parameter estimates. The standard errors used to conduct significance tests of parameter estimates are simply the square roots of the diagonal elements of the matrix [1]. The standard errors in SAS would then report the hierarchical arrangement of the sampled endemic transmission-oriented parameter estimators by statistical significance employing the "covariance parameter estimates" table based on the diagonal entries. If an empirical standard error matrix with the EMPIRICAL option in the PROC GLIMMIX statement is then used for regressing the seasonal-sampled vector arthropod-related spatial data feature attributes, a likelihood-based sandwich estimator could also be computed based on the subject-specific gradients of the Laplace or quadrature approximation. The sandwich estimator would then simply replace H^{-1} for determining the predictive convergence in the endemic transmission-oriented risk model. Thereafter, to compute the standard errors and prediction standard errors of linear combinations of β and γ, in the seasonal arthropod-related endemic transmission-oriented predictive multivariate risk model, PROC GLIMMIX would employ an approximate prediction variance matrix. This probabilistic estimation matrix would then have to be formulated for $[\hat{\beta}, \hat{\gamma}]'$ from $P = \begin{bmatrix} H^{-1} & H^{-1}(\frac{\partial \gamma}{\partial [\beta, \theta]}) \\ (\frac{\partial \gamma}{\partial [\beta, \theta]})H^{-1} & \Gamma^{-1} + (\frac{\partial \gamma}{\partial [\beta, \theta]})H^{-1}(\frac{\partial \gamma}{\partial [\beta, \theta]}) \end{bmatrix}$, where Γ would be the second derivative matrix from the γ suboptimization that maximizes $f(y, \beta, \theta, \gamma)$ for the values of β and θ. As such, the prediction variance sub-matrix for the random effects would be based on approximated conditional MSE of the predictions. Note that even in the normal linear mixed model, the approximate conditional prediction standard errors would not be identical to the prediction standard errors obtained by the inversion of the mixed multivariate vector arthropod-related endemic transmission-oriented predictive risk model regression-based equations.

In terms of quantitating conditional fit and output statistics for optimal seasonal multivariate predictive arthropod-related model estimation however, the parameters of a mixed model must employ Laplace approximation or quadrature in GLIMMIX. By doing so, the procedure could then display fit statistics related to the marginal distribution as well as the conditional distribution $p(y|\hat{\gamma}, \hat{\beta}, \hat{\phi})$. For ODS purposes, the name of the "conditional fit statistics" table would be "CondFitStatistics" (www.sas.edu) as the marginal likelihood would be approximated numerically for deriving robust seasonally predictive arthropod-related endemic transmission-oriented multivariate model residual forecasts parsimoniously. These estimators would be based on the marginal distribution which currently is not available for robust seasonal predictive arthropod-related risk modeling. Instead the generalized Pearson Chi-square statistic in PROC GLIMMIX would utilize the "fit statistics" table, to report the Pearson statistic of the conditional distribution in the "conditional fit statistics" table. By doing so, the unavailability of the marginal distribution in the seasonal multivariate vector arthropod-related residual forecasts targeting the statistically significant endemic transmission-oriented explanatory covariate coefficients would affect the set of output statistics produced with METHOD=LAPLACE and METHOD=QUAD. Unfortunately, output statistics and statistical graphics that depend on the marginal variance of the sampled arthropod-related data would not be available with these estimation methods.

Alternatively, if a user-defined variance function for a seasonal multivariate vector arthropod-related predictive risk model is constructed in PROC GLIMMIX, the conditional distribution of the sampled endemic transmission-oriented data would then be unknown. Laplace or quadrature estimation would then not be possible. When specifying a variance function with METHOD=LAPLACE or METHOD=QUAD, the procedure would assume that the conditional distribution in the predictive endemic transmission model is normal. For example, consider the following statements to fit a mixed model to count data:

```
Proc glimmix method=Laplace;
    Class sub;
    _variance_ = _phi_ * _mu_;
    Model count = x / s link=log;
    Random int / sub=sub;
Run;
```

Thereafter, the variance function and the link could then suggest an overdispersed Poisson model. However, the Poisson distribution would not be able to accommodate the extra scale parameter in this situation. Instead, the GLIMMIX procedure would need to fit a mixed model with random intercepts, log link function and variance function $\phi\mu$, assuming that the endemic transmission-oriented count variable is normally distributed. This variable would then be given by the random effects in the predictive risk model. Fortunately, a new bias-corrected estimator is available in PROC NL MIXED for generating seasonal robust multivariate vector arthropod-related endemic transmission-oriented predictive risk model residual forecasts. This would be the COVTEST statement in PROC NL MIXED which may enable likelihood-based inference about the seasonal covariance vector arthropod-related endemic transmission-oriented larval habitat parameters.

As such, in this research, we investigated "default" priors for the variance components in PROC NL MIXED for constructing multiple robust seasonal predictive multivariate riverine larval habitat distribution models of *Similium damnosum s.l.,* using multiple georeferenced endemic transmission oriented observational explanatory covariate coefficients spatiotemporally sampled in Nabere village, in Burkina Faso. We propose a new prior distribution for classical (non-hierarchical) multivariate vector arthropod-related endemic transmission-oriented seasonal predictive regression models. Our model was constructed by first scaling all non-binary variables to have mean 0 and standard deviation 0.5. Thereafter, independent student-t prior distributions on the seasonal-sampled coefficients were determined. As a default choice, we employed the Cauchy distribution with center 0 and scale 2.5, which

was then subsequently employed for determining a longer-tailed version of the distribution by assuming one-half additional success and one-half additional failure in a logistic regression model. Cross-validation on a corpus of datasets then revealed a Cauchy class of prior distributions, which in this research outperformed existing implementations of Gaussian and Laplace priors. The Cauchy distribution has no moment generating function, but it is closely related to the Poisson kernel, which is the fundamental solution for the Laplace equation in the upper half-plane [8].

We then utilized this prior distribution as a default choice in our construction phase of our seasonal multivariate predictive *S. damnosum s.l.* riverine larval habitat endemic transmission oriented risk model. Our assumption was that there would be a complete separation in the regression model for applying more shrinkage to higher-order interactions for generating robust residual forecasts targeting important explanatory covariates. We assumed that this separation may then be useful in routine S. *damnosum s.l.* larval habitat predictive data risk analysis as well as in automated procedures, such as chained equations for missing-data imputation. We implemented the procedure to fit various GLMs in SAS/GIS employing the student-t prior distribution by incorporating an approximate EM algorithm along with iteratively weighted least squares. In statistics, an EM algorithm is an iterative method for finding ML or maximum a posteriori (MAP) estimates of parameters in statistical models, especially where the model depends on unobserved latent variables [1].

We then employed the PRIOR statement in an SAS/GIS environment, which enabled us to carry out a robust sampling-based on a seasonal Bayesian hierarchical regression-based analysis in PROC NL MIXED. The analysis produced an SAS/GIS dataset containing a pseudo-random sample from the joint posterior density of the variance components and the other regressed parameters, which was then included in a mixed spatiotemporal model. The posterior analysis was then performed after all other NL MIXED computation procedures were completed.

An independence chain algorithm was then employed to generate a pseudo-random proposal from a convenient base distribution of the

regressed seasonal-sampled georeferenced *S. damnosum s.l.* riverine larval habitat explanatory endemic transmission-oriented explanatory covariate coefficient estimates. These coefficient estimates were chosen to be as close as possible to the posterior. The proposal was then retained in the sample with a probability proportional to the ratio of the seasonal-sampled larval habitat weights constructed by quantitating the ratio of the true posterior to the base density. In order to better approximate the marginal posterior density of the variance components in the endemic transmission-oriented predictive risk model forecasts, PROC NL MIXED transformed the sampled explanatory covariate coefficients by using MIVQUE (0) equations. In order to better approximate the marginal posterior density of the variance components, PROC NL MIXED then transformed the fixed-effects parameters. By doing so, uncertainty estimates and other non-normal error probabilities were analytically integrated out of the joint posterior leaving the marginal posterior density of the variance components in the regressed MIVQUE (0) equation outputs. Thereafter, the NL MIXED procedure performed the selected transformation employing the PTRANS option and specifying the TDATA= option. The density of the transformed S. *damnosum s.l.* riverine larval habitat parameters were then approximated by a product of inverted gamma densities.

The density of the transformed *S. damnosum s.l.* riverine larval habitat estimators were then also approximated by product of inverted gamma densities. In probability theory and statistics, the inverse gamma distribution is a two-parameter family of continuous probability distributions on the positive real line, which coincidentally is the distribution of the reciprocal of multivariate regressed georeferenced predictor variables distributed according to the gamma distribution [8]. Gamma distribution is a two-parameter scale parameter θ, which has a shape parameter k, therefore, if k is an integer, the distribution would represents an Erlang distribution whereby, the sum of k independent exponentially distributed random variables, each have a mean of θ which would be mathematically equivalent to a rate parameter of θ^{-1}. The gamma distribution is frequently a probability model in time series-dependent arthropod-related infectious disease larval habitat data

analyses [2]. In this research, the seasonal predictive *S. damnosum s.l.* riverine larval habitat modeled the gamma distribution, which was also the maximum probability distribution for a random variate X in the model residuals forecast for which $E(X) = a$ was fixed and greater than zero. The residual forecasts where then quantitated by $E(\ln(X)) = \psi(a)$, which in this research was fixed when $(\psi(\cdot))$ was the digamma function employing the logarithmic derivative of the gamma function.

The digamma function is defined as the logarithmic derivative of the gamma function: $\psi'(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ [8]. Meanwhile, the gamma is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers [1]. That is, if n is a positive integer in a seasonal predictive *S. damnosum s.l.* larval habitat endemic transmission-oriented multivariate risk model, the gamma function would then be defined for all complex numbers except the non-positive integers. For complex sampled seasonal values with a positive real part in a predictive risk model, this function would then be defined via an improper integral that converges. Theoretically, this integral function could be extended by analytic continuation to all the seasonal-sampled *S. damnosum s.l.* endemic transmission-oriented explanatory covariate coefficient values in a predictive risk model except the non-positive integers of the function would have simple poles thereby, yielding a meromorphic function (i.e., gamma function). Although the gamma function is a component which has been employed in various probability-distribution functions, including combinatorics, it has not been applied extensively to seasonal vector arthropod-related endemic transmission-oriented predictive multivariate risk modeling.

Additionally, in this research, to determine the seasonal-sampled *S. damnosum s.l.* riverine larval habitat estimators for the inverted gamma densities, PROC NL MIXED evaluated the logarithm of the posterior density over a digitized ArcGIS QuickBird–derived digitized grid-based residual algorithmic matrix using multiple sampled georeferenced points. QuickBird satellite (www.digitalglobe.com) is an excellent source of environmental data useful for analyses of changes in

land usage, agricultural and forested riverine climates related to seasonal *S. damnosum s.l.* riverine larval habitats and their associated spatial feature attributes [2].

PROC NL MIXED then performed a linear regression of these values on the logarithm of the inverted gamma density values rendered from the distribution of the remote-sampled endemic transmission- oriented explanatory predictor covariate coefficient estimates. This lead to a *d*-dimensional analogue of the inverse-gamma-normal conjugate prior for normalized sampling in one dimension. Popular Bayesian generalized hierarchical model builds upon the linear regression employing conjugate priors since they can be used to infer properties other than the mean function, such as the conditional variance or response quantiles [1]. A conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior [8]. As such, we employed the conjugate prior to generate a robust seasonal Bayesian nonparametric mixture model to quantitate the sampled observational explanatory predictor variables to determine both the number and form of the local mean function in the sampled *S. damnosum s.l.* riverine larval habitat data.

Further, in this research, we also decomposed the Wishart probability distribution of the sample covariance matrix generated in PROC NL MIXED by employing quantitated probability distributions of eigenvalues and eigenvectors in order to determine optimal non-linear *S. damnosum s.l.* parameter estimators. Wishart distribution is a generalization to multiple dimensions of the Chi-squared distribution, or, in the case of non-integer degrees of freedom (df), of the gamma distribution whereby, any of a family of error probability distributions can be defined over symmetric, nonnegative-definite matrix-valued random variables [2]. Our assumption was that these distributions may be of great importance in the estimation of covariance matrices in time series-dependent robust predictive autoregressive multivariate *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented risk modeling. In Bayesian inference, the Wishart distribution is of particular importance, as it is the conjugate prior of the inverse of the covariance matrix (i.e., the precision matrix) of a multivariate normal

distribution [10]. An important use of the Wishart distribution for spatiotemporal vector arthropod-related data analyses is a conjugate prior for multivariate normal sampling [1, 2].

We then constructed a semiparametric spatial filtering approach in SAS/GIS (http://ftp.sas.com) to deal explicitly with the uncertainty (e.g., residual heteroscedascity) in the spatiotemporal *S. damnosum s.l.* riverine larval habitat distribution model forecasts by reducing the number of parameter estimators. This was done by using spatially lagged autoregressive models. SAS/GIS is an interactive geographic information system (GIS) within the SAS system that has an open data model, meaning the information stored in both the attribute and spatial datasets. Further, SAS spatial datasets must conform to the topological rules outlined and published by the American Society for Photogrammetry and Remote Sensing and the American Congress on Surveying and Mapping (www.esri.com). These rules include topological completeness and topological geometric consistency. Spatial files failing to meet the topological criteria can cause errors, alerting a user that quality control is necessary, if spatial analysis is to be conducted. PROC GIS creates and maintains spatial datasets for use in SAS (www.esri.com).

In this research, PROC MAPIMPORT imported ESRI shape-files into SAS. A user interface exists in the solutions menu in an interactive GIS window for supporting predictive multivariate seasonal vector arthropod-related platforms [2]. Unfortunately, this interface was not very intuitive for producing sophisticated seasonal multivariate vector arthropod-related endemic transmission-oriented predictive risk maps. As such, in this research, the riverine larval habitat modeling was performed employing using SAS/GRAPH. This program created various different types of maps in PROC GMAP: two dimensional choropleth maps and three-dimensional block, prism, and surface maps. SAS and ESRI have partnered to create a bi-directional bridge between SAS data and analytical tools and the ESRI mapping environment. This bridge has been implemented recently by the U.S. Bureau of the Census to create school district demographics (www.esri.com).

In this research, we assumed that the observed spatial patterns in the response variable rendered from a dataset of regressed georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented explanatory covariate coefficients could be decomposed into three statistically independent components including: (a) a systematic spatial trend component that could be specified by a parsimonious set of exogeneous variables; (b) a stochastic signal that reflected either an underlying spatial process and/or a set of missing exogeneous factors with an inherent spatial pattern; and (c) independent white-noise disturbances. An assumption in this research was that a specific subset of eigenvectors from a transformed spatial link matrix could be used to capture dependencies among the disturbances of the serially correlated predictive autoregressive *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented risk model parameter estimators. As such, the residual estimates from the off-diagonal elements of a covariance matrix were then generated from a spatial filter eigenevector analysis prior by exporting the sampled data into a Bayesian probabilistic estimation framework using WinBUGS® (www.mrcbsu.cam.ac.uk/bugs/).

Further, in this research, Bayesian statistics in WinBUGSio® and spatial filter eigenvectors from SAS/GIS® (http://ftp.sas.com) were employed for constructing robust endemic multivariate transmission-seasonal field-operational risk maps. We developed the framework for a remote habitat-based surveillance system thereafter employing PROC NL MIXED, SAS/GIS, WinBUGSio and satellite-derived landscape-oriented models. We then decomposed the Wishart probability distribution of the sample covariance matrix by employing models generated in PROC NL MIXED and SAS/GIS into probability distributions of eigenvalues and eigenvectors in order to calculate multiple seasonal Bayesian error estimation models using the sampled georeferenced *S. damnosum s.l.* endemic transmission-oriented explanatory covariates. Generalizations of the multivariate inverse gamma densities include Wishart distributions [8]. An important use of the Wishart distribution is as a conjugate prior for multivariate normal sampling [1]. In this research, we also generated a semiparametric

spatial filtering approach in SAS/GIS to deal explicitly with uncertainty in the *S. damnosum s.l.* larval habitat distribution model by reducing the number of data parameters using spatially lagged autoregressive models and simultaneous autoregressive spatial models. Residual estimates from the off-diagonal elements of a covariance matrix were then generated from the spatial filter eigenevector analysis prior to exporting the sampled data into a Bayesian estimation matrix by also using WinBUGS®.

We employed WinBUGSio, an SAS macro program, which facilitates remote execution of WinBUGS from within SAS. This is an SAS macro, which does the data handling and input/output from WinBUGS® via SAS®. In this research, the program produced a column format data file based on the spatiotemporally-sampled georeferenced *S. damnosum s.l.* riverine larval habitat mutlivariate endemic transmission oriented explanatory predictor covariate coefficient estimates and also wrote a list format data file for constants. The macro program then wrote a script file to the WinBUGS directory referencing the appropriate datafile, model file; init file, and log file names. The script then ran WinBUGS in batch mode which read in the node statistics block from the log file. Although in this research, there was a requirement to specify the input and output file names and directory path as well as the statistics to be monitored in WinBUGS, the code checked for optimal convergence diagnostics within the geodatabase. We also checked for non-normal distributed random errors in the regression equation. Methods for regressing spatiotemporal vector arthropod-related larval habitat data rely on the assumption of normality and the use of linear estimation methods (e.g., least squares) to make probabilistic inferences [2].

In our Bayesian model, we considered the inverse-Wishart distribution, which is a proper conjugate prior for an unknown covariance matrix in a multivariate normal model. Some specific analytical results for the inverse-Wishart have been derived; for example, the marginal distribution of a diagonal block submatrix of draws from an inverse-Wishart distribution [7]. Various alternatives to the inverse-Wishart have been proposed but even fewer analytical results are known for these families, making it even more challenging to understand precisely the

properties of such distributions. Consequently, analytical understanding of these distributions falls short of providing a full understanding of the inverse-Wishart distribution for any type of predictive arthropod-related risk modeling.

Thereafter, in this research, we performed a non-smooth optimization by stabilizing the steepest descent in the riverine risk model by exploiting gradient and subgradient information in the model. In this paper, we investigated the behaviour of quasi-Newton (i.e., variable metric) methods, specifically, the well-known Broyden-Fletcher-Goldfarb-Shannon (BFGS) method, to minimize non-smooth functions, both convex and nonconvex in the *S. damnosum s.l.* riverine larval habitat model. In optimization, quasi-Newton methods (a special case of variable metric methods) are algorithms for finding local maxima and minima of functions [8]. Quasi-Newton methods are also a generalization of the secant method to find the root of the first derivative for multi-dimensional problem [1]. In multiple dimensions, the secant equation is underdetermined for Quasi-Newton methods, which are often used to find the stationary point of a function where the gradient is 0.

The behaviour of quasi-Newton methods on non-smooth functions has received no attention for predictive multivariate seasonal vector arthropod-related endemic transmission-oriented risk modeling. While any locally Lipschitz non-smooth function f can be viewed as a limit of increasingly ill-conditioned differentiable functions [1] via "mollifiers", for example, most of them may have no consequence for the algorithm's asymptotic convergence behaviour when f is not differentiable at its minimizer. These are commonly seen in smooth functions with special properties employed in distribution theory to create sequences of smooth functions approximating non-smooth (i.e., generalized) functions, via convolution however, they have never been applied for seasonal predictive vector arthropod-related risk modeling.

Lipschitz continuity is an important concept in mathematical analysis. In modern variational analysis, it has been generalized for set-valued mappings. Among many extensions, the pseudo-Lipschitzian property has been well recognized as a natural and useful one. It is now

called by different names, such as the Aubin property or the Lipschitz-like property [10]. The concept has been used extensively in the study of sensitivity analysis of optimization problems and variational inequalities. It also plays an important role on developing generalized differentiation calculi for non-smooth functions and set-valued mappings. According to [9] for a set-valued mapping $F : Rm \rightarrow\rightarrow Rn$, the Aubin property around $(\bar{}x, \bar{}y) \in \text{gph } F := \{(x, y) \in Rm \times Rn \mid y \in F(x)\}$, if there exist neighbourhoods V of $\bar{}x$, W of $\bar{}y$, and a constant $\ell \geq 0$ such that $F(x) \cap W \subseteq F(u) + \ell kx - ukIB$ for all $x, u \in V$.

Intuitively, then given a function in a seasonal multivariate predictive *S. damnosum s.l.* larval habitat-related endemic transmission model, which is rather irregular, by convolving it with a mollifier, the function may get "mollified", that is, its sharp features (e.g., georeferenced) spatial feature attributes will be smoothed, while still remaining close to the original non-smooth (i.e., generalized) function. Further, when applied to a wide variety of non-smooth, locally Lipschitz functions, not necessarily convex, the BFGS method may be very effective for seasonally quantitating seasonal-sampled S. *damnosum s.l.* riverine larval habitat data using the gradient difference information to update an inverse Hessian approximation (*Hk*). The Hessian is updated by analyzing successive gradient vectors [1]. Although quasi-Newton methods differ in how they constrain the solution, they may be easily quantized by adding a simple low-rank update to the current estimate of the Hessian for effectively targeting endemic transmission oriented S. *damnosum s.l* related explanatory covariate coefficients.

The most common quasi-Newton algorithms for predictive risk modeling are currently the SR1 formula (for symmetric rank one), the BHHH method, the widespread BFGS method (suggested independently by Broyden, Fletcher, Goldfarb, and Shannon, in 1970), and its low-memory extension, LBFGS. The Broyden's class is a linear combination of the DFP and BFGS methods. Unfortunately, the SR1 formula will not guarantee the update matrix to maintain positive-definiteness in a seasonal multivariate predictive *S. damnosum s.l.* larval habitat related endemic transmission-oriented risk model. The Broyden's method

however may update the predictive matrices to be symmetric, which may be then employed to find the root of a general system of equations in the risk model rather than the gradient for updating the Jacobian or the Hessian. One of the chief advantages of quasi-Newton methods over Newton's method is that the Hessian matrix (or, in the case of quasi-Newton methods, its approximation), B does not need to be inverted [8]. Newton's method and its derivatives, such as interior point methods, may require the Hessian to be inverted, in a robust seasonal predictive *S. damnosum s.l.* riverine larval habitat model, however, it may be implemented by solving a system of linear equations. In contrast, quasi-Newton methods may generate an estimate of B^{-1} directly for determining statistically significant riverine larval habitat model residual forecasts.

We considered three classes of error distributions in our *S. damnosum s.l.* larval habitat riverine larval habitat multivariate seasonal endemic trasnmission oriented predictive risk model as a useful alternative to the normal distribution: t-distributions, generalized error distributions, and Tukey's contaminated normal distribution. We assumed that these distributions were robust in the sense that any outliers in the seasonal-sampled *S. damnosum s.l.* riverine larval habitat empirical ecological dataset logically would have less of an effect on the estimated mean (i.e., regression) function than on the residual forecasts targeting the statistically significant endemic transmission-oriented based normal/non-normal distributions. At an intuitive non-mathematical level, the use of heavy tailed distributions in a seasonal infectious disease model allows for a small number of large error residuals to be quantitated [8]. Further, we assumed since a normal distribution would force the predicted residuals to be within a few standard deviations of the mean, the *S. damnosum s.l.* larval habitat outliers could adversely affect the estimated mean, biasing the estimate in the predictive model by greatly inflating the estimated standard error of the mean. This would cause loss of predictability power by increasing the width of confidence intervals in the risk model [2].

Therefore, our hypothesis in this paper was that PROC NL MIXED SAS/GIS, WinBUGSio, and QuickBird-derived regression-based risk models could account for any multivariate predicted variability in seasonal sampled larval habitat productivity of *S. damnosum s.l.* sampled in a riverine ecosystem study site. Our objectives in this research were to: (1) construct robust ecological-weighted regression equations and global autocorrelation statistics to identify linear and non-linear based explanatory predictors associated to productive habitats; (2) quantitate latent autocorrelation uncertainty coefficients in the model residual forecasts using a covariance matrix rendered from an eigenfunction decomposition spatial filter algorithm; and (3) generate inverse-Wishart priors within a Bayesian probabilistic estimation framework for forecasting the regression -based distribution of multiple field and remote-sampled endemic transmission-oriented *S. damnosum s.l.* riverine larval habitat explanatory covariate coefficients spatiotemporally sampled in Nabere village in Burkina Faso. This research constitutes the first attempt to extend a conventional spatial autoregressive *S. damnosum s.l.* riverine larval habitat distribution model by using a regression analyses in PROC NL MIXED, an eigenvector model formulation in SAS/GIS and Bayesian regression probabilistic estimation analyses in an SAS®macro WinBUGSio environment for identifying observational endemic transmission explanatory covariates coefficients. Although the discussion is centered on onchocerciasis vectors, the framework and derived guidelines in this research may be applicable to integrated control programs for other vector species and arthropod-borne infectious diseases.

2. Material and Methods

2.1. Study site

This research was conducted in Nabere village in Burkina Faso. The study site area consists of extensive plains, low hills and valleys, high savannas, and a desert area in the north. Nabere village has three distinct seasons: warm and dry (November-March), hot and dry (March-May), and hot and wet (June-October). Annual rainfall varies from about 1,000mm to less than 250mm. The terrain is mostly flat to dissected

undulating plains with hills at the peripheral of the study site. Most of the study site region lies on a savanna plateau, 198-305m above sea level, with fields, brush, and scattered trees.

2.2. A PROC cluster-based classification

A 5km buffer was placed around known 5 georeferenced epidemiological breeding sites using the Landsat Thematic data in ArcGIS 10.1®. We then created buffer polygons around the georeferenced spatiotemporal-sampled *S. damnosum s.l* riverine larval habitats input features of the 5 sample sites. FLEXIBLE|FLE in SAS 9.2® (Carey, North Carolina) was then used to request the flexible-beta method. The clustering methods in SAS presently include average linkage, the centroid method, complete linkage, density linkage (including Wong's hybrid and *k*-th nearest-neighbour methods), ML for mixtures of spherical multivariate normal distributions with equal variances but possibly unequal mixing proportions, the flexible-beta method, McQuitty's similarity analysis, the median method, single linkage, two-stage density linkage, and Ward's minimum-variance method (http://ftp.sas.com). PROC CLUSTER then displayed the table of eigenvalues of the covariance matrix for canonical variables. Generally, in a PROC CLUSTER table, output the first two columns list each eigenvalue and the difference between the eigenvalue and its successor, while the last two columns display the individual and cumulative proportion of variation associated with each eigenvalue (www.sas.edu). In our model, the squared multiple correlations, R^2, was the proportion of variance accounted for by the *S. damnosum s.l.* riverine larval habitat clusters. The approximate expected value of R^2 was then given in the column labelled "ERSQ". The next three columns displaced the values of the cubic clustering criterion (CCC), pseudo F (PSF), and t2 (PST2) statistics. These statistics were useful in determining the number of *S. damnosum s.l.* clusters in the data. A method of judging the number of clusters in a dataset in PROC CLUSTER is to look at the pseudo F statistic (PSF) (www.sas.edu).

The CLUSTER procedure hierarchically clustered the seasonal-sampled *S. damnosum s.l.* observations using SAS data. The data was

then analyzed by using the *S. damnosum s.l.* sampled geocoordinates and squared Euclidean distances measurements within a flexible-beta method in PROC CLUSTER.

The PROC CLUSTER statement started the procedure which specified a clustering method based on Annual Biting Rates (ABR) values and then optionally specified each georeferenced varying and constant explanatory *S. damnosum s.l.* endemic transmission oriented covariate within the clusters. The ABR then estimated the number of infective bites a person receives during one-year period. It was calculated by multiplying the ABR with the proportion of infective mosquitoes in the biting population. The agglomerative hierarchical clustering procedure then used the sampled observations to create multiple clusters based on the georeferenced spatiotemporal reverine larval habitat data in a georeferenced riverine larval habitat cluster by itself. Clusters were then merged to form a new cluster that replaced the two old clusters. Merging of the two closest clusters was repeated until only one cluster was left (Figure 1).

Figure 1. Hierarchical ABR-stratified cluster-based analyses using georeferenced *S. damnosum s.l.* riverine larval habitat sites.

In this research, beta was set at –100 for both cluster-based analyses in SAS PROC CLUSTER. The flexible-beta method began by specifying METHOD=FLEXIBLE. PROC CLUSTER then created an output *S. damnosum s.l.* riverine larval habitat empirical dataset that revealed a cluster hierarchy of the sampled data feature attributes based on the ABR values. Since in this research, the georeferenced parameter estimators were deemed to be equally important, we employed the STD option in PROC CLUSTER to standardize the cluster-based varying and constant endemic transmission oriented explanatory predictor covariate coefficients to mean 0 and standard deviation. Covariates with large variances tend to have more effect on the resulting spatial clusters than variables with small variances but, if all coefficients are considered equally important in a model, the STD option in PROC CLUSTER can standardize the sampled variable (www.sas.com). In this research, the STDIZE procedure standardized the spatiotemporal-sampled *S. damnosum s.l.* larval habitat parameter estimators in the SAS dataset by subtracting the georeferenced larval habitat location measures and then dividing them by a scale measure. Finally, a unique identifier was incorporated for each cluster.

In order to reduce the likelihood of chaining among the ABR cluster-based varying and constant *S. damnosum s.l.* endemic transmission oriented explanatory predictor covariates, a partition that best represented the sampled parameter estimates was identified. This was performed by finding the intersection between a manageable number of interpretable cluster-based varying and constant explanatory covariates and then quantitating them with large jumps in the normalized *S. damnosum s.l.* larval habitat distance measurements, in PROC CLUSTER. The cluster-based parameter estimates were then plotted against GIS-based Euclidean distance measurements which revealed a clear flattening of the curve in the overlain data indicating that adequate separation of the cluster-based varying and constant endemic transmission-oriented explanatory covariate coefficients could not be achieved beyond a specific georeferenced larval habitat point. The number of spatiotemporal-sampled *S. damnosum s.l.* riverine larval habitat clusters in the data was also determined by preliminary

evaluations with varying numbers of cluster solutions aimed at avoiding trivial error by plotting the georeferenced larval habitat data in discriminant function space and by seeking adequate separation among group centroids. In order to compute meaningful standardized rates, the individual sampled riverine larval habitat observational predictors were then aggregated geographically into high-low ABR stratified clusters. The final model revealed that among the highest density ABR-based clusters was the Nabere epidemiological study site.

2.3. Environmental data analyses

We generated univariate statistics and regression models by using the data stored. We also generated a misspecification term for constructing a spatial autoregressive model. Multiple data layers were then created using different coded values for the various *S. damnosum s.l.* riverine larval habitat spatial feature attributes. Distance measurements and canopy coverage measures were then calculated by using the QuickBird data and the field-sampling information (Table 1).

Table 1. Environmental sampled within cluster based varying and constant covariates of *S. damnosum s.l.* in the Nabere study site as entered in SAS®

Variable	Description	Units
GCP	Ground control points	Decimal-degrees
FLOW	Flowing water	Presence or absence
DISCAP	Distance from capture point	Meters
ELEV	Elevation	Meters
AQVEG	Aquatic vegetation	Percentage
HGVEG	Hanging vegetation	Percentage
FLVEG	Floating vegetation	Percentage
MMB	Man-made barriers	Type (e.g., damns)
DISHAB	Distance between habitats	Meters

2.4. Regression analyses

The relationship between sampled immature *S. damnosum s.l.* habitat and each individual sampled endemic transmission-oriented

predictive risk-related-explanatory covariate was then investigated by single variable regression analysis by using PROC MIXED. Since parasite prevalence data are binomial fractions, a regression model was used, as is standard practice for the analysis of such data [1, 2]. A Poisson regression analyses was then used to determine the relationship between *S. damnosum s.l.* habitat larval count data and the sampled habitat characteristics.

The regression analyses assumed independent counts (i.e., *Ni*), taken at sampled habitat locations $i = 1, 2, \ldots, n$. The *S. damnosum s.l.* habitat larval counts were described by a set of variables denoted by matrix $\mathbf{X}_i$, where a $1 \times p$ vector of covariate coefficient indicator values for a sampled larval habitat location i. The expected value of these data was given by $\mu_i(\mathbf{X}_i) = n_i(\mathbf{X}_i)\exp(\mathbf{X}_i\beta)$, where β was the vector of non-redundant parameters in the *S. damnosum s.l.* larval habitat model where the Poisson rates were given by $\lambda_i(\mathbf{X}_i) = \mu_i(\mathbf{X}_i)/n_i(\mathbf{X}_i)$. In this research, rates parameter $\lambda_i(\mathbf{X}_i)$ was both the mean and the variance of the Poisson distribution for each sampled *S. damnosum s.l.* larval habitat location i. The dependent variable was total seasonal sampled larval density count. The Poisson regression model assumed that the sampled *S. damnosum s.l.* larval habitat data was equally dispersed, that is, that the conditional variance equaled the condition mean. The procedure used ML estimation to find the regression coefficients. The data was then log-transformed before analyses to normalize the distribution and minimize standard error.

There was considerable overdispersion in the regression-based model; thus, we used a negative binomial model to determine parameters associated to the seasonal-sampled *S. damnosum s.l.* larval habitat data. Overdispersion is often encountered when fitting very simple parametric models such as those based on the Poisson distribution [10]. If over-dispersion is a feature in a vector larval habitat distribution model, an alternative model with additional free parameters may provide a better fit [9]. In this research, a Poisson mixture model with a negative binomial distribution was used, where the mean of the Poisson distribution was

itself a random variable drawn from the gamma distribution; thereby, introducing an additional free parameter in the spatiotemporal-sampled *S. damnosum s.l.* larval habitat distribution model. The family of negative binomial distributions is a two-parameter family, which uses several parameterizations for treating overdispersion data [1]. The Poisson distribution has one free parameter and does not allow for the variance to be adjusted independently of the mean [8]. In this research, a parameterization technique was employed such that two variables p and r with $0 < p < 1$ and $r > 0$. Under this parameterization, the probability mass function (pmf) of the ecological-sampled *S. damnosum s.l.* riverine larval habitat predictor variables with a NegBin (r, p) distribution took the following form: for $k = f(k; r, p) = \binom{k + r - 1}{k} \cdot p^r \cdot (1 - p)^k 0, 1, 2,$ where $\binom{k + r - 1}{k} = \frac{\Gamma(k + r)}{k!\,\Gamma(r)} = (-1)^k \cdot \binom{-r}{k}$ and $\Gamma(r) = (r - 1)!$. We also used an alternative parameterization for the sampled *S. damnosum s.l.* larval habitat data using the mean $\lambda : \lambda = r \cdot \left(p^{-1} - 1\right) p = \frac{r}{r + \lambda}$ and the mass function then became: $g(k) = \frac{\lambda^k}{k!} \cdot \frac{\Gamma(r + k)}{\Gamma(r)(r + \lambda)^k} \cdot \frac{1}{\left(1 + \frac{\lambda}{r}\right)^r}$, where λ and r were the sampled parameters. Under this parameterization, we were able to generate: $\lim_{r \to \infty} g(k) = \frac{\lambda^k}{k!} \cdot 1 \cdot \frac{1}{\exp(\lambda)}$, which was the mass function of a Poisson-distributed random variable with Poisson rate λ. In other words, the alternatively parameterized negative binomial distribution generated from the regressed riverine larval habitat explanatory covariates converged to the Poisson distribution, and r controlled the deviation from the Poisson. This made the negative binomial habitat model suitable as a robust alternative to the Poisson model for modeling the *S. damnosum s.l.* seasonal-sampled endemic transmission-oriented risk-related-explanatory covariates.

In this research, the negative binomial distribution of the sampled explanatory covariates arose as a continuous mixture of Poisson distributions, where the mixing distribution of the Poisson rate was a gamma distribution. The mass function of the negative binomial distribution of the predictor variables then was written as

$$f(k) = \int_0^\infty Poisson(k|\lambda) \cdot Gamma(\lambda|r, (1-p)/p) d\lambda$$

$$= \int_0^\infty \frac{\lambda^k}{k!} \exp(-\lambda) \cdot \frac{\lambda^{r-1} \exp(-\lambda p/(1-p))}{\Gamma(r)((1-p)/p)^r} d\lambda$$

$$= \frac{1}{k!\,\Gamma(r)} p^r \frac{1}{(1-p)^r} \int_0^\infty \lambda^{(r+k)-1} \exp(-\lambda/(1-p)) d\lambda$$

$$= \frac{1}{k!\,\Gamma(r)} p^r \frac{1}{(1-p)^r} (1-p)^{r+k} \Gamma(r+k)$$

$$= \frac{\Gamma(r+k)}{k!\,\Gamma(r)} p^r (1-p)^k.$$

2.5. Spatial analyses

Initially, a misspecification perspective for the estimation models was generated assuming that the spatiotemporal *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk-based-model using $y = X\beta + \varepsilon^*$ (i.e., regression equation) for quantitating autocorrelated disturbances ε^*. In this research the autocorrelation coefficients were decomposed into a white-noise component, ε, and a set of unspecified and/or misspecified models that had the structure $y = XB + \underbrace{E\gamma + \varepsilon}_{=\varepsilon^*}$. White noise is a univariate or multivariate discrete-time stochastic process, whose terms are independent and identically distributed (i.i.d.) with a zero mean [2]. In this research, the misspecification term was $E\gamma$. Quantification of the topographic patterns rendered from the distribution of the sampled georeferenced *S. damnosum s.l.* larval habitat

explanatory covariates was required to describe independent key dimensions of the underlying spatial processes in the habitat data for defining a spatial pattern in the misspecification term.

A spatial autoregressive model was then generated that employed a sampled *S. damnosum s.l.* riverine larval habitat variable, **Y**, as a function of nearby sampled habitat **Y** endemic transmission oriented explanatory covariate coefficient indicator value *I* (i.e., an autoregressive response) and/or the residuals of **Y** as a function of nearby sampled habitat **Y** residuals (i.e., an SAR or spatial error specification). For larval habitat modeling, the SAR model furnishes an alternative specification that frequently is written in terms of matrix **W** [9]. As such, in this research the spatial covariance of the sampled dataset was a function of the matrix $(\mathbf{I} - \rho\mathbf{C}\mathbf{D}^{-1})(\mathbf{I} - \rho\mathbf{D}^{-1}\mathbf{C}) = (\mathbf{I} - \rho\mathbf{W}^{T})(\mathbf{I} - \rho\mathbf{W})$, where T denoted matrix transpose. The resulting matrix was symmetric and was considered a second-order specification as it included the product of two spatial structure matrices (i.e., $\mathbf{W}^{T}\mathbf{W}$). This matrix restricted positive values of the autoregressive parameter to the more intuitively interpretable range of $0 \le \hat{\rho} \le 1$.

In this research, distance between sampled larval habitats was defined in terms of an *n*-by-*n* geographic weights matrix, **C**, whose c_{ij} values were, 1 if the sampled *S. damnosum s.l.* riverine larval habitat locations i and j were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum rendered **C1**, where **1** was an *n*-by-**1** vector of ones, which subsequently converted the regression-based matrix to matrix **W**. The resulting SAR model specification, with no sampled larval habitat covariates present (i.e., the pure spatial autoregression specification), took on the following form: $\mathbf{Y} = \mu(1 - \rho)\mathbf{1} + \rho\mathbf{W}\mathbf{Y} + \varepsilon$, where μ was the scalar conditional mean of **Y**, and ε was an *n*-by-1 error vector whose parameters were statistically i.i.d. normally random variates. The spatial covariance matrix for analyzing the sampled georeferenced explanatory covariates coefficients was thereafter $E[(\mathbf{Y} - \mu\mathbf{1})'(\mathbf{Y} - \mu\mathbf{1})] = \Sigma = [(\mathbf{I} - \rho\mathbf{W}')(\mathbf{I} - \rho\mathbf{W})]^{-1}\sigma^{2}$,

where $E(\bullet)$ denoted the calculus of expectations, $\mathbf{I}$ was the n-by-n identity matrix denoting the matrix transpose operation, and σ^2 was the error variance.

We then employed a Hessian matrix. As in Newton's method, we used a second order approximation to find the minimum of a function $f(x)$ in the predictive *S. damnosum s.l.* larval habitat model. We then employed a Taylor series of $f(x)$ to iterate $f(x_k + \Delta x) \approx f(x_k) + \nabla f(x_k)^T \Delta x + \frac{1}{2}\Delta x^T B \Delta x$, where (∇f) was the gradient and B was an approximation to the Hessian matrix. In mathematics, the Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a function [1]. In this research, this matix described the local curvature of a function of the sampled variables. Given the function $f(x_1, x_2, \ldots, x_n)$, if all second partial derivatives of f exist and are continuous over the domain of the function, then the Hessian matrix of f is $H(f)_{ij}(\boldsymbol{x}) = D_i D_j f(\boldsymbol{x})$, where $x = (x_1, x_2, \ldots, x_n)$ and D_i is the differentiation operator with respect to the i-th argument [9]. The matrix rendered in SAS was

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The Hessian matrix is related to the Jacobian matrix by $H(f)(x) = J(\nabla f)(x)$ [9]. The determinant of the above matrix was Hessian [1]. Hessian matrices are used in large-scale optimization problems within Newton-type methods because they are the coefficient of the quadratic term of a local Taylor expansion of a function [7]. Thus, in the predictive multivariate larval habitat risk model, $y = \frac{1}{2} f(\boldsymbol{x}^T + \Delta \boldsymbol{x}) \approx f(\boldsymbol{x}) + J(\boldsymbol{x})\Delta \boldsymbol{x} + \Delta \boldsymbol{x}$, $H(\mathbf{x})\Delta \mathbf{x}$, where J was the Jacobian matrix, was a vector

(i.e., the gradient) for the scalar-valued functions. The full Hessian matrix can be difficult to compute in practice; in such situations, quasi-Newton algorithms have been developed that use approximations to the Hessian [9].

In vector calculus, the Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function [9]. Specifically, suppose $F : \mathbb{R}^n \to \mathbb{R}^m$ is a function which takes as input real n-tuples and produces as output real m-tuples. Such a function would be given by m real-valued component functions, $F_1(x_1, \ldots, x_n), \ldots, F_m(x_1, \ldots, x_n)$ [2]. In this research, the partial derivatives of all the functions with respect to the variables $x_1, \ldots, x_n$ was organized in an m-by-n matrix whereby, the Jacobian matrix J of F, was

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}.$$

The matrix entries were then functions of $x_1, \ldots, x_n$, which were then denoted by $J_F(x_1, \ldots, x_n)$ and $\frac{\partial(F_1, \ldots, F_m)}{\partial(x_1, \ldots, x_n)}$.

The Jacobian matrix revealed that the function F was differentiable at a sampled *S. damnosum s.l.* riverine larval habitat point $p = (x_1, \ldots, x_n)$, which was slightly stronger than the derivative of F at p in the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ represented by the matrix $J_F(x_1, \ldots, x_n)$. This linear transformation was the best linear approximation of the function F near the sampled larval habitat point p. Since $m = n$ in our Jacobian matrix, it was a square matrix and its determinant was function of $x_1, \ldots, x_n$, which was the Jacobian determinant of F. The matrix about the local behaviour of F was then thought of as a local expansion factor for multiple volumes. It was used for performing variable substitutions in the riverine larval habitat multi-variable integrals.

The gradient of the approximation with respect to Δx, we noted was then $\nabla f(x_k + \Delta x) \approx \nabla f(x_k) + B\Delta x$. Thereafter, by setting this gradient to zero, the Newton step: $\Delta x = -B^{-1}\nabla f(x_k)$ was rendered. The Hessian approximation B was chosen to satisfy $\nabla f(x_k + \Delta x) = \nabla f(x_k) + B\Delta x$ (i.e., secant equation). We then noted that in more than one dimension, B was underdetermined in the matrix. In one dimension, solving for B and applying the Newton's step with the updated value was equivalent to the secant method. The various quasi-Newton methods differ in their choice of the solution to the secant equation (in one dimension, all the variants are equivalent) [11]. Most methods for predictive risk modeling seek a symmetric solution $(B^T = B)$ whereby, the variants are motivated by finding an update B_{k+1} that is as close as possible to B_k, that is, $B_{k+1} = \arg\min_B \|B - B_k\|_V$, where V is a positive definite matrix. In this research this solution was defined by the norm. We used the Broyden's method. In numerical analysis, Broyden's method is a quasi-Newton method for the root-finding algorithm in k variables [8].

Newton's method for solving the equation $f(x) = 0$ then was applied to the Jacobian matrix, J, at every iteration. However, computing this Jacobian was a difficult operation. The idea behind Broyden's method is to compute the whole Jacobian only at the first iteration, and to do a rank-one update at the other iteration [9].

An approximate initial value of $B_0 = I * x$ was then used to achieve rapid convergence. The unknown x_k was then updated applying the Newton's step, which was calculated by using the current approximate Hessian matrix $B_k\Delta x_k = -\alpha_k B_k^{-1}\nabla f(x_k)$, with α chosen to satisfy the Wolfe conditions; $x_{k+1} = x_k + \Delta x_k$. The gradient was then computed at a new *S. damnosum s.l.* larval habitat point $\nabla f(x_{k+1})$, and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, which in this research was used to update the approximate Hessian B_{k+1}, or directly its inverse $H_{k+1} = B_{k+1}^{-1}$ using the Sherman-Morrison formula. The Sherman-Morrison formula computed the inverse of the sum of an invertible matrix A and the outer product, uv^T, of vectors u and v.

Next, an autoregressive model specification was generated. The model was written as: $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$, where $\varphi_1, \ldots, \varphi_p$ were the field and remote-sampled *S. damnosum s.l.* larval habitat parameters of the model, c was a constant, and ε_t was the white noise. When coupled with regression and the normal probability model, an autoregressive specification results in a covariation term characterizing spatial autocorrelation by denoting the autoregressive parameter with ρ at a conditional autoregressive covariance specification [1]. In this research this specification involved the matrix $(\mathbf{I} - \rho\mathbf{C})$, where $\mathbf{I}$ was an n-by-n identity matrix. In an autoregressive expression; however, the response variable is on the left-side of the equation, while the spatial lagged version of this variable is on the right side [9]. Therefore, one of the main objectives in this research was to bring the spatially unlagged endogeneous variable, y, exclusively on the left-hand side of the regression equation in order to decorrelate the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related explanatory covariate coefficients. In this research, this was accomplished by expanding the matrix term: $(I - \rho V)^{-1} = \sum_{k=0}^{\infty} \rho^k V^k$, as an infinite power series, which was feasible under the assumption that the underlying spatial process in the sampled ecological datasets was stationary. The simultaneous autoregressive error model was then rewritten as $y - \rho Vy = X\beta - \rho VX\beta + \varepsilon$. Substituting this transformation rendered

$$y = (I - \rho V)^{-1}[X\beta - \rho V(X\beta) + \varepsilon],$$

$$y = \sum_{k=0}^{\infty} \rho^k V^k (X\beta - \rho VX\beta + \varepsilon),$$

$$y = \sum_{k=0}^{\infty} \rho^k V^k X\beta - \sum_{k=0}^{\infty} \rho^{k+1} V^{k+1} (X\beta) + \sum_{k=0}^{\infty} \rho^k V^k \varepsilon,$$

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty}\rho^k V^k X\beta - \sum_{k=1}^{\infty}\rho^k V^k (X\beta)}_{=0} + \sum_{k=0}^{\infty}\rho^k V^k \varepsilon,$$

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty}\rho^k V^k \varepsilon}_{misspecification\ term} + \varepsilon.$$

The misspecification term $\sum \rho^k V^k \varepsilon$ $(k = 1, \ldots, \infty)$ meanwhile remained uncorrelated with the exogeneous variable, X, as the standard OLS assumption of the disturbances, ε, were uncorrelated with the larval habitat predictor variables generated from the parameter estimates. The spatial lag model was then expressed as $(I - \rho V)y = X\beta + \varepsilon$. Substituting the transformation generated: $y = \sum_{k=0}^{\infty}\rho^k V^k (X\beta + \varepsilon)$ and $y = X\beta + \underbrace{\sum_{k=1}^{\infty}\rho^k V^k (X\beta + \varepsilon)}_{misspecification\ term} + \varepsilon$. The misspecification term $\sum \rho^k V^k (X\beta + \varepsilon)$ $(k = 1, \ldots, \infty)$ included the exogeneous variables X. Consequently, the exogeneous variables were correlated with the misspecification term. Under this condition, standard OLS results for the basic regression model $y = X\beta + \varepsilon^*$ generated from the sampled georeferenced larval habitat endemic transmission oriented explanatory covariates provided biased estimates $\hat{\beta}$ of the underlying regression parameters β.

2.6. Eigenvector analyses

The correlation, or lack thereof, between the exogeneous variables and the misspecification terms of both *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk models were then used to design spatial proxy variables, so that the properties of either model could be satisfied. We considered two different projection matrices, $M_{(1)} \equiv I - 1\left(1^T 1\right)^{-1} 1^T$ and $M_{(X)} \equiv I - X\left(X^T X\right)^{-1} X^T$. The projection

matrix $M_{(1)}$ is a special case of the more general projection matrix $M_{(X)}$ [11]. The general projection matrix $M_{(X)}$ included a constant unity vector 1 and multiple additional exogeneous variables. The set of eigenvectors $\{e_1, \ldots, e_n\}_{SAR}$ was then extracted from the quadratic form

$$\{e_1, \ldots, e_n\}_{SAR} \equiv evec\left[M_{(X)} \frac{1}{2}\left(V + V^T\right) M_{(X)} \right], \tag{2.1}$$

which in this research was designed orthogonal to the exogeneous variable *X*. The projection matrix $M_{(X)}$ imposed this constraint. In contrast, the set of eigenvectors $\{e_1, \ldots, e_n\}_{Lag}$ was extracted from

$$\{e_1, \ldots, e_n\}_{Lag} \equiv evec\left[M_{(1)} \frac{1}{2}\left(V + V^T\right) M_{(1)} \right]. \tag{2.2}$$

These two different sets of eigenvectors established a basis for constructing a spatiotemporal *S. damnosum s.l.* larval habitat regression-based distribution model. Both expressions were then solely defined in terms of exogeneous information. This model feature enabled us to also use the eigenvector spatial filtering approach for predictions of the endogeneous variable *y*. The associated sets of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}_{Lag}$ and $\{\lambda_1, \ldots, \lambda_n\}_{SAR}$, with $\lambda_i \geq \lambda_{i+1}$, range were then used for properly standardizing adjacent link matrices *V* that were related to irregular spatial tessellations generated from the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission oriented explanatory covariate coefficients.

The components of each eigenvector, e_i, where then mapped onto an underlying spatial tessellation which exhibited a distinctive topographic pattern ranging from PSA (i.e., similar values of log-transformed larval count data aggregating in geospace) for $\lambda_i > E(I)$, to NSA (i.e., dissimilar log-values aggregating in geospace) for $\lambda_i > E(I)$. Each eigenvector was mapped where $E(I)$ was the expected value of Moran's *I* under the assumption of (a) spatial independences and (b) as outputs from related projection matrices $M_{(1)}$ or $M_{(X)}$, respectively. We noted that

the associated Moran's I autocorrelation coefficient, of each eigenvector e_i generated was equal to its associated eigenvalue $\lambda_i = \left[e_i^T \left(V + V^T\right) e_i\right] / \left(2 e_i^T e_i\right)$, but only if V was scaled to satisfy $\left[1^T \left(V + V^T\right) 1\right] / 2 = n$. Moran's autocorrelation is often denoted as I is an extension of Pearson's product moment correlation coefficient which can be used to measure the amount of autocorrelation in an ecological-sampled datasets of seasonal sampled multivariate vector arthropod-related habitat weighted estimators [2]. In previous research, Jacob et al. [11] employed the Pearson's correlation coefficient for spatially summarizing autocovariance between two sampled malarial mosquito *Anopheles. arabiensis* larval habitat predictor variables to define the covariance of any sampled explanatory covariate coefficients divided by the product of their standard deviations using $\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$. The formula defined the sampled larval habitat population correlation coefficient. In this research, substituting estimates of the covariances and variances provided the sample correlation coefficient denoted by $r : r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$. An equivalent expression rendered the correlation coefficient as the mean of the products of the standard scores. Based on paired *S. damnosum s.l.* larval habitat data (i.e., X_i, Y_i), the sample Pearson correlation coefficient was $r - \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s_X} \right) \left(\frac{Y_i - \overline{Y}}{s_X} \right)$, where $\frac{X_i - \overline{X}}{s_X}$, $\overline{X}$, and s_X was the standard score, sample mean, and sample standard deviation, respectively. The eigenvectors yielded distinct risk-oriented map pattern descriptions of latent spatial autocorrelation in the larval habitat data. This was interpreted as synthetic map variables that represented specific natures (i.e., positive or negative) and degrees (i.e., negligible, weak, moderate, and strong) of potential spatial autocorrelation. For the immature Similium, two counteracting spatial autocorrelation effects were conceptualized (i.e., common factors leading to PSA and competitive factors leading to NSA materializing) at the same time, with a possible net effect being global detection of near-zero spatial autocorrelation. If a

parsimonious set of eigenvectors is to be selected for eigenvectors depicting near-zero spatial autocorrelation should be avoided, since such a set of latent vectors associated with a matrix equation fail to capture any geographic information [9].

In this research, the eigenvector spatial filtering approach added a minimally sufficient set of eigenvectors as proxy-variables to the set of linear predictors, in the *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk-related model by inducing mutual independence in the sampled parameter estimator datasets. The regression residuals represented spatially independent variable components. The spatial pattern in the eigenvectors was synthetic. In our model, positive global autocorrelation in the local patterns of the larval habitat parameters exhibited only positive local autocorrelation and vice versa for negative global autocorrelation. The eigenvectors e_i and e_j within each set of eigenvectors, were mutually orthogonal as the symmetry transformation $\frac{1}{2}\left(V + V^T\right)$ was a quadratic form as revealed in Equations (2.1) and (2.2).

As mentioned previously, the eigenvectors of specification (2.1) were orthogonal to the exogeneous variables X of the regression model constructed employing the sampled georeferenced *S. damnosum s.l.* larval habitat explanatory covariates; whereas, the eigenvectors of specification (2.2) were orthogonal only to the constant unity vector 1 in X. This orthogonality had implications for modeling the spatial misspecification terms in the larval habitat model and allowed us to link each collection of eigenvectors to its specific autoregressive model by letting E_{SAR} be a matrix whose vectors were subsets of $\{e_1, \ldots, e_n\}_{SAR}$. A linear combination of this subset was then approximated by employing the misspecification term of the simultaneous autoregressive version of the *S. damnosum s.l.* riverine larval habitat distribution endemic transmission-oriented predictive risk-related-model

$$(\text{i.e., } E_{SAR}\gamma \approx \sum_{k=1}^{\infty} \rho^k V^k \varepsilon). \tag{2.3}$$

The linear combination $E_{SAR}\gamma$ remained orthogonal to exogeneous variables X and, consequently, the estimated field and remote-sampled explanatory predictor variables $\hat{\beta}$ were unbiased. Furthermore, as a property of the OLS estimator, the estimated term $E_{SAR}\gamma$ was also orthogonal to the residuals $\hat{\varepsilon}$. The model $y = X\hat{\beta} + E_{SAR}\hat{\gamma} + \hat{\varepsilon}$ then decomposed the endogeneous variable y into a systematic trend component, a stochastic signal component and some white-noise residuals. The term $E_{SAR}\hat{\gamma}$ removed variance inflation in the MSE term attributable to spatial autocorrelation.

Alternatively, for the spatial lag model (2.3), E_{Lag} was a matrix of those eigenvectors that were a subset of $\{e_1, \ldots, e_n\}_{Lag}$. The approximation of the misspecification term became $E_{Lag}\gamma \approx \sum_{k=0}^{\infty} \rho^k V^k (X\beta + \varepsilon)$. Since $E_{Lag}\gamma$ is correlated with the exogeneous variables X, its incorporation into the *S. damnosum s.l.* larval habitat predictive risk model corrected the bias of estimated plain OLS parameters $\hat{\beta}$ in the spatial lag analyses. The model $y = X\hat{\beta} + E_{Lag}\hat{\gamma} + \hat{\varepsilon}$, generated from the sampled georeferenced riverine larval habitat explanatory covariates was a decomposition of the spatial lag model into a systematic trend component, a stochastic signal component and some white-noise residuals. However, for the sampled *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk model, the trend and the stochastic signal were no longer uncorrelated and the MSE was deflated.

The set of eigenvectors $\{e_1, \ldots, e_n\}_{Lag}$ of the spatial lag model (2.3) was then calculated independently of the exogeneous variables X. In this research, this calculation was dependent on the underlying spatial link matrix V. We found that, this filtering approach was more adaptable to an explanatory specification search of relevant exogeneous variables and spatial predictions with changing predictor variable values in the model. In contrast, for the simultaneous autoregressive *S. damnosum s.l.*

riverine larval habitat model (2.2), the eigenvectors $\{e_1, \ldots, e_n\}_{SAR}$ depended through the projection of $M_{(X)}$, on the exogeneous variables X. Thus, any change in the underlying model structure required a recalculation of the eigenvectors for generating the tessellations. Spatial filtering of either the spatial lag model or the simultaneous autoregressive model with a common factor constraint thereafter, only required identification of only one set of selected eigenvectors, namely, E_{SAR} or E_{Lag}, respectively [2]. The relevant set of eigenvectors was then applied simultaneously to all the field and remote-sampled georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk-related explanatory covariates in both models. For the generic autoregressive model (2.1); however, spatial filtering was applied individually to each sampled explanatory predictor covariate coefficient. The generic specification of autoregressive spatial models then associated a specific spatial lag factor with the endogeneous y variable and other specific spatial lag factors for each additional exogeneous variable [9]. In this research, we used the eigenvectors $\{e_1, \ldots, e_n\}_{Lag}$ to filter spatial autocorrelation in the generic autoregressive *S. damnosum s.l.* larval habitat model employing each sampled parameter estimator.

The next step was to identify suitable and parsimonious subsets of eigenvectors E_{SAR} or E_{Lag} from either sampled *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk model specification (2.1) or (2.2). A particular subset of eigenvectors is suitable, if the residuals $\hat{\varepsilon}$ of the resulting spatially filtered model become stochastically independent with respect to the underlying sampled habitat spatial structure V [11]. In addition, parsimony in the model estimation in this research was defined as the smallest possible subset of eigenvectors which led to the spatial independence in the *S. damnosum s.l.* larval habitat distribution model residuals being identified. The spatial patterns of different eigenvectors expressed independent and filter spatial autocorrelation of the regression model residuals which in this research was explicitly formalized by a georeferenced vector. This methodology has been used for extrapolating

georeferenced explanatory covariates associated to prolific larval habitats. For example, in Jacob et al. [11], an eigenfunction decomposition algorithm was used along with a forward stepwise regression to add eigenvectors to regression-based malarial mosquito *Anopheles gambiae s.l.* and West Nile Virus mosquito *Culex quinquefasciatus* aquatic larval habitat models for targeting prolific habitats in Gulu, Uganda until the spatial autocorrelation in the resulting residuals $\hat{e}$ dropped below a critical level. The measures of clustering of *An. gambiae s.l.* and *Cx. quinquefasciatus* aquatic larval habitats were then reported. In this research, results from SAS PROC GENMOD for all the *S. damnosum s.l.* riverine larval habitat model autocorrelation eigenvectors were selected by the stepwise negative binomial regression procedure. Positive and negative spatial autocorrelation spatial filter component pseudo-R^2 values were then reported by using GLMM estimation results from SAS PROC NLMIXED.

2.7. Bayesian matrix

In this research, we began with a "prior distribution", which was based on the relative likelihoods of the sampled georeferenced *S. damnosum s.l.* larval habitat "Bayesianized" endemic transmission-oriented predictive risk-related explanatory covariates. In practice, it is common to assume a uniform distribution over the appropriate range of values for the prior distribution [9]. We calculated the likelihood of the observed distribution as a function of the parameter estimator values, which multiplied this likelihood function by the prior distribution which was then normalized to obtain a unit probability over all possible values (i.e., posterior distribution). The mode of the distribution was then the parameter estimate and "probability intervals". These intervals represented the Bayesian analogue of confidence intervals in the regression-based distribution model.

In our Bayesian formulation, the specification of the *S. damnosum s.l.* riverine larval habitat model was completed by assigning priors to all unknown parameters. We used our dataset of spatiotemporal-sampled observations where each x_i for $i = 1, \ldots, n$ was assumed to be distributed according to some distribution $p(x_i|\theta)$. In this research, θ

was a parameter that was unknown and had to be inferred from the data. Our Bayesian procedure began by assuming that θ was distributed according to some prior distribution $p(\theta|\alpha)$, where the parameter α was a hyperparameter. The joint probability of the sampled *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related data was then determined by using: $p(\mathbf{X}|\theta) = p(x_1, \ldots x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$; whereby, $p(\mathbf{X}|\theta, \alpha) = p(\mathbf{X}|\theta)$ and $p(x_i|\theta, \alpha) = p(x_i|\theta)$ was conditionally independent of the hyperparameters. Bayesian inference then determined the posterior distribution of the parameter $p(\theta|\mathbf{X}, \alpha)$ by using

$$
\begin{aligned}
p(\theta|\mathbf{X}, \alpha) &= \frac{p(\theta, \mathbf{X}, \alpha)}{p(\mathbf{X}, \alpha)} \\
&= \frac{p(\theta, \mathbf{X}, \alpha)}{\int_{\theta} p(\theta, \mathbf{X}, \alpha) d\theta} \\
&= \frac{p(\mathbf{X}|\theta, \alpha)p(\theta|\alpha)}{\int_{\theta} p(\mathbf{X}|\theta, \alpha)p(\theta|\alpha) d\theta} \\
&= \frac{p(\mathbf{X}|\theta)p(\theta|\alpha)}{\int_{\theta} p(\mathbf{X}|\theta)p(\theta|\alpha) d\theta} \\
&= \frac{\left[\prod_{i=1}^{n} p(x_i|\theta)\right] p(\theta|\alpha)}{\int_{\theta} \left[\prod_{i=1}^{n} p(x_i|\theta)\right] p(\theta|\alpha) d\theta}.
\end{aligned}
$$

For the fixed regression parameters, a suitable choice was the diffuse prior, i.e., $p(\gamma)$ *const*., but a weakly informative Gaussian prior was also possible. A second-order Gaussian random walk prior was then employed to allow enough flexibility while penalizing abrupt changes in the function.

Initially, we generated heterogeneous random walks in one dimension employing the sampled riverine larval habitat model parameter estimators. We simulated a random walk by first drawing a random number out of a uniform distribution that determined the propagation direction according to the transition probabilities, and then drew a random time out of the relevant different jumping time probability density functions (JT-PDF). We noted that the interval in our sampled data was discrete. In a discrete system, the connections are among adjacent states, while the dynamics are either Markovian, semi-Markovian, or even non-Markovian depending on the model heterogeneous random walks in 1-D. This systems have jump probabilities that depend on the location in the system, and/or different jumping time (JT) probability density functions (PDFs) that depend on the location in the system [1]. Known important results in simple systems include a symmetric Markovian random walk, the Green's function (i.e., PDF) of the walker for occupying state i, which is commonly Gaussian and has a variance that scales like time [11]. This results in a system with discrete time and space, and also in a system continuous time and space. We used a completely heterogeneous semi-Markovian random walk in a discrete system of $L(>1)$, where the Green's function was found in Laplace space for quantifying the *S. damnosum s.l.* riverine larval habitat spatiotemporal-sampled data.

In this research, the Laplace transform of a function was defined with, $\bar{f}(s) = \int_0^\infty e^{-st} f(t)dt$. Thereafter, the system was defined through the jumping time (JT) PDFs: $\psi_{ij}(t)$ connecting states i with state j. The jump was from state i. The solution was based on the path representation of the Green's function, calculated by including all the path PDFs of the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission oriented explanatory covariates by using $\overline{G}_{ij}(s) = \overline{\Gamma}_{ij}(s)\dfrac{\overline{\Phi}(s,\ \widetilde{L})}{\overline{\Phi}(s,\ L)}\overline{\Psi}_i(s)$. We then derived $\overline{\Psi}_i(s) = \sum_j \overline{\Psi}_{ij}(s)$ and $\overline{\Psi}_{ij}(s) = \dfrac{1-\overline{\psi}_{ij}(s)}{s}$ from

$\overline{G}_{ij}(s) = \overline{\Gamma}_{ij}(s)\dfrac{\overline{\Phi}(s, \overline{L})}{\overline{\Phi}(s, L)}\overline{\Psi}_i(s)$. Thereafter we, used these equations to generate the following equations in SASmacro WinBUGSio : $\overline{\Gamma}_{ij}(s) = \prod_{c=0}^{i\mp1}\psi_{c\pm1c}(s)$, $\overline{\Phi}(s, L) = 1 + \sum_{c=1}^{[L/2]}(-1)^c\overline{h}(s, c; L)$ with $\overline{h}(s, i; L) = \prod_{c=1}^{i}\sum_{k_c=2+k_{c-1}}^{L-1-2(i-c)}\bar{f}_{k_c}(s)$ and $\bar{f}_{k_j}(s) = \psi_{k_jk_j+1}(s)\psi_{k_j+1k_j}(s)$. By doing so, $L = I$, $\overline{\Phi}(s; L) = 1$. The symbol $[L/2]$ then appeared in the upper bound in the $\sum$ in the regression-based equations which was the floor operation. This essentially entailed rounding off the sampled georeferenced explanatory predictor covariates towards zero. In this research, the factor $\Phi(s; \widetilde{L})_{I)}$ had the same form as in $\overline{\Phi}(s; L)$ which was calculated on a lattice $\widetilde{L}$. Lattice $\widetilde{L}$ was constructed from the original lattice generated from the larval habitat dataset by taking out from it the states i and j and the states between them, and then connecting the two fragments. When each fragment is a single state, $\overline{\Phi}(s; \widetilde{L}) = 1$ [9].

Next, a random walk having a step size that varied according to a normal distribution of the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related explanatory covariates was determined. The Black-Scholes formula then employed a Gaussian random walk as an underlying assumption. Here, the step size was the inverse cumulative normal distribution $\Phi^{-1}(z, \mu, \sigma)$, where $0 \le z \le 1$ were the sampled larval habitat density count values μ and σ were the mean and standard deviations of the normal distribution, respectively. The Black-Scholes equation we used was a partial differential equation. The equation was $\dfrac{\partial V}{\partial t} + \dfrac{1}{2}\sigma^2S^2\dfrac{\partial^2 V}{\partial S^2} + rS\dfrac{\partial V}{\partial S} - rV = 0$. We noticed that our data followed a classic geometric Brownian motion (GBM). That is, $\dfrac{dS}{S} = \mu dt + \sigma dW$, where W was Brownian motion. The GBM (i.e., exponential Brownian motion) is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion [11]. Brownian motion is the presumably random

drifting in a mathematical model used to describe random movements [1]. Note that W, and consequently, its infinitesimal increment dW, were the only source of uncertainty in our model. Intuitively, $W(t)$ was then a process that fluctuated up and down in such a random way that its expected change over any time interval was 0. The variance T in $W(t)$ was then used for constructing Gaussian random walks of spatiotemporal larval habitat distribution models. Therefore in the model, the expected value of μ was dt with a variance of $\sigma^2 dt$ [1].

In this research, the stochastic process S_t generated from the field and remote sampled *S. damnosum s.l.* larval habitat parameter estimates was said to follow a GBM which then satisfied the following stochastic differential equation (SDE): $dS_t = \mu S_t dt + \sigma S_t dW_t$, where W_t was a Brownian motion and μ (the percentage drift), and σ (i.e., the percentage volatility) were constants. For an arbitrary initial value S_0, the above SDE had the analytic solution under Itō's interpretation: $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$, which was for any value of t, a log normally distributed random variable with expected value $\mathbb{E}(S_t) = S_0 e^{\mu t}$ and variance $\mathrm{Var}(S_t) = S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)$. The correctness of this solution was checked by using Itō's lemma.

In its simplest form, Itō's lemma states the following: for an Itō drift-diffusion process $dX_t = \mu_t dt + \sigma_t dB_t$ and any twice differentiable function $f(t, x)$ of two real variables t and x, one has $df(t, X_t) = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{\sigma_t^2}{2}\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma_t \frac{\partial f}{\partial x} dB_t$. This immediately implies that $f(t, X)$ is itself an Itō drift-diffusion process. In higher dimensions, Ito's lemma states $df(t, X_t) = \dot{f}_t(t, X_t)dt + \nabla_{X_t}^T f \cdot dX_t + \frac{1}{2} dX_t^T \cdot \nabla_{X_t}^2 f \cdot dX_t$, where $X_t = (X_{t,1}, X_{t,2}, \ldots, X_{t,n})^T$ is a vector of Itō processes, $\dot{f}_t(t, X)$ is

the partial differential w.r.t. t, $\nabla_X^T f$ is the gradient of f w.r.t. X, and $\nabla_X^2 f$ is the Hessian matrix of f w.r.t. X. More generally, the above formula also holds for any continuous d-dimensional semimartingale $X = (X^1, X^2, \ldots, X^d)$, and twice continuously differentiable and real valued function f on $\mathbf{R}^d$. Some authors prefer to present the formula in another form with cross-variation shown explicitly as follows: $f(X)$ is a semimartingale satisfying $df(X_t) = \sum_{i=1}^{d} f_i(X_t) dX_t^i + \frac{1}{2} \sum_{i,j=1}^{d} f_{i,j}(X_t) d[X^i, X^j]_t$. In this expression, the term f_i represents the partial derivative of $f(x)$ with respect to x^i, and $[X^i, X^j]$ the quadratic covariation process of X^i and X^j.

Finally, for the spatial components, *VI*, a Markov random field (MRF) prior was assigned. The conditional distribution of *VI* gave an adjacent explanatory covariate, *VJ*, which was a univariate normal distribution with mean equal to the average *VJ* coefficient values of *VI*'s neighbouring sampled site variance equal to τ_v^2 divided by the number of total adjacent sampled *S. damnosum s.l.* larval habitat endemic transmission oriented estimators. This lead to a joint density of the form: $p\left(\nu \middle| \tau_\nu^2\right) \alpha \exp\left(-\frac{\tau_\nu^2}{2} \sum_{i-j} (\nu_i - \nu_j)^2 \right)$, where $i \sim j$ denoted sampled habitats i adjacent to j, when the estimates *VI* and *VJ* in the adjacent sampled habitats were similar. The degree of uncertainty in the spatiotemporal-sampled data was then determined by the unknown precision parameter τ_ν^2.

By writing $f_j = Z_j \beta_j$, $h = Z_k \beta_k$, $u = Z_l \beta_l$, and $\nu = Z_m \beta_m$ for a well-defined design matrix Z, a vector of regression parameters β, with all different priors, was expressed in a general Gaussian form by using the expression: $p\left(\beta_j \middle| \tau_\nu^2\right) \alpha \exp\left(-\frac{1}{2\tau_j^2} \beta_j' K_j \beta_j \right)$ with an appropriate penalty

matrix K_j. The model structure was dependent on the sampled georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission oriented explanatory covariates and smoothness of the function. In most cases, K_j is rank deficient and, hence, the prior for β_j is improper [7]. For the variances τ_j^2 inverse Gamma priors $IG(a_j, b_j)$ was assumed with hyperparameters a_j, b_j chosen such that this prior was weakly informative.

The Bayesian framework in this research was then defined by conditional probabilities constructed from spatiotemporal-sampled *S. damnosum s.l.* larval habitat data. The observation nodes in the Bayesian estimation model was denoted by a vector $x = (x_1, x_2, \ldots, x_N)$, and the set of states of the observation node x_j were generated from the sampled data which was represented by $x_j \in \{1, 2, \ldots, Y_j\}$. In this research, the hidden nodes were denoted by $z_k \in \{1, 2, \ldots, T_k\}$. The probability that the state of the hidden node z_k was $i, 1 \le i \le T_k$, was expressed as $a_{(k,i)} \coloneqq P(z_k = i)$. Because $\{a_{(k,i)}, i = 1, 2, \ldots, T_k\}$ is a probability distribution, $\sum_{i=1}^{T_k} a_{(k,i)} = 1$ it holds for $k = 1, 2, \ldots, K$ [9]. The conditional probability in our model was based on the *j*-th riverine larval habitat observation when x_j while $l, (1 \le l \le Y_j)$ were based on the condition that the states of hidden nodes were $[Z = (Z]_1, Z_2, \ldots, Z_K)$, generated by $b_{(j,l|z)} \coloneqq P(x_j = l|z)$. We then defined $a \coloneqq \{a_{(k,i)}\}, b \coloneqq \{b_{(j,l|z)}\}$. Thereafter we employed $\omega = \{a, b\}$ as the set of all the spatiotemporal-sampled *S. damnosum s.l.* larval habitat parameters estimators spatiotemporally sampled at the Nabere study site. Then the joint probability that the states of observation nodes were $x = (x_1, x_2, \ldots, x_N)$ and the states of hidden nodes were quantitated by $[z = (z]_1, z_2, \ldots, z_K)$, based on $P(x, z|\omega) = \prod_{k=1}^{K} a_{(k,z_k)} \prod_{j=1}^{N} b_{(j,x_j|z)}$.

Thereafter, the marginal probability that the states of observation *S. damnosum s.l.* riverine larval habitat nodes were x was generated by using $P(x, z|\omega) = \sum_{z} P(x, z|\omega) = \left\{\prod_{k=1}^{K} \sum_{z_k=1}^{T_k} \square\right\} \prod_{k=1}^{K} a_{(k, z_k)} \prod_{j=1}^{N} b_{(j, x_j|z)}$. We employed the notation $\sum_{z} = \left\{\prod_{k=1}^{K} \sum_{z_k=1}^{T_k} \square\right\} := \sum_{z_1=1}^{T_1} \sum_{z_2=1}^{T_2} \cdots \sum_{z_K=1}^{T_K} \square$ for the summation over all states of the hidden nodes. We assumed the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related explanatory covariates $X_n = \{X_1, X_1, \ldots, X_n\}$ were independently and identically taken from the true distribution $p_0(x)$. In Bayesian learning, the prior distribution $\varphi(\omega)$ on the parameter ω is set [7]. In this research, the posterior distribution $p(\omega|X^n)$ was computed from the *S. damnosum s.l.* larval habitat dataset and the prior by $p(\omega|X^n) = \frac{1}{Z(X^n)} \exp(-nH_n(\omega))\varphi(\omega)$, which then generated the expression $H_{\downarrow}n(\omega) = 1/n \sum_{\downarrow}(i = 1)^{\uparrow} n \equiv \log(po\{X_{\downarrow}i\))/(p\{X_{\downarrow}i\)|\omega))$, and $Z(X^n)$ (i.e., the normalization constant). The Bayesian predictive distribution $p(x|X^n)$ was provided by averaging the model over the posterior distribution as follows: $p(x|X^n) = \int [p(x|\omega)p(\omega|\square]X^n)d\omega$. The Bayesian stochastic complexity $F(Xn)$ was then defined by $F(X^n) = -\log Z(X^n)$, which was then used as a criterion by which the model selected the hyperparameters in the prior. We then let $E_{x^n}[\cdot]$ be the expectation over all the sampled *S. damnosum s.l.* larval habitat parameters. The Bayesian stochastic complexity had the following asymptotic form: $E_{\downarrow}(X^{\uparrow}n)[F(X^{\uparrow}n) \approx \lambda \log n - (m-1)\log\log n + O(1)$, where λ and m were the larval habitats geosampled and their endemic transmission oriented explanatory covariate coefficient indicator count values, respectively. In regular models, 2λ is equal to the number of parameters and $m = 1$, while in non-identifiable models, 2λ is not larger than the number of parameters and $m \geq 1$ [9]. However, our Bayesian framework

constructed using field and remote-sampled habitat data required integration over the posterior distribution, which could not be performed analytically [1].

Thus, in this research, we let $\{X^n, Z^n\}$ be the spatiotemporal-sampled *S. damnosum s.l.* larval habitat parameter estimators corresponding to the hidden error variables in the equation $Z^n = \{Z_1, Z_1, \ldots, Z_n\}$. The variational framework approximated the Bayesian posterior $p(Z^n, \omega|X^n)$ based on hidden variables and the variational posterior $q(Z^n, \omega|X^n)$, which was factorized using $q(Z^n, \omega|X^n) = Q(Z^n|X^n)r(\omega|X^n)$, where $Q(Z^n|X^n)$ and $r(\omega|X^n)$ were posteriors based on the inconspicuous error coefficients and the sampled data, respectively. The variational posterior $q(Z^n, \omega|X^n)$ was then chosen to minimize the functional $\overline{F}_{[q]}$ defined by $\overline{F}_{[q]} = \sum_{Z^n} \int \frac{q(Z^n, \omega|X^n)\log\left(q(Z^n, \omega|X^n)po(X^n)\right)}{p(X^n, Z^n, \omega)} d\omega$, which in this research was then further defined by $s = F(X^{\uparrow}n) + K(q(Z^{\uparrow}n, \omega|X^{\uparrow}n)||p(Z^{\uparrow}n, \omega|X^{\uparrow}n))$ using the sampled parameters, where $K(q(Z^{\uparrow}n, \omega|X^{\uparrow}n)||p(Z^{\uparrow}n, \omega|X^{\uparrow}n))$, where the true Bayesian posterior was $p(Z^n, \omega|X^n)$ and the variational posterior was $q(Z^n, \omega|X^n)$. This led to the functional $\overline{F}_{[q]}$ being minimized under the constraint and then the variation posteriors, $r(\omega|X^n)$ and $Q(Z^n|X^n)$, was computed by using the equation $r(\omega|X^n) = \frac{1}{C_r}\varphi(\omega)\exp < \log p(X^n|Z^n|\omega) > Q$ and $Q(Z^n|X^n) = \frac{1}{C_Q}\exp < \log p(X^n, Z^n|\omega) > r$, where C_r and C_Q were the normalization constants. It is important to note that these equations gave only necessary conditions for the functional $\overline{F[q]}$ to be minimized in the *S. damnosum s.l.* riverine larval habitat distribution model. The variational posteriors were computed by an iterative algorithm. We

defined the variational stochastic complexity $\overline{F(X^n)}$ by the minimum value of the functional $\overline{F[q]}$, which was $\overline{F}(X^n) = \min_{r,Q} \overline{F[q]}$, based on the difference between $\overline{F(X^n)}$ and the Bayesian stochastic complexity $F(X^n)$.

We then generated variational posterior for the estimation matrix. We assumed that the prior distribution $\varphi(\omega)$ of the *S. damnosum s.l.* larval habitat parameters $\omega = \{a, b\}$ was the conjugate prior distribution. Then $\varphi(\omega)$ was given by $\varphi(a_k) = \frac{\Gamma(T_k \phi_0)}{\Gamma(\phi_0)^{T_k}} \prod_{z_k=1}^{T_k} a_{(k,z_k)}^{\phi_0 - 1}$, $k = 1, 2, \ldots, K$, using the sampled georeferenced explanatory riverine larval endemic transmission oriented covariate coefficient estimates which were given by $\varphi(b_{(j_k|z)}) = \frac{\Gamma(Y_j \xi_0)}{\Gamma(\xi_0)^{Y_j}} \prod_{x_j=1}^{Y_j} b_{(j,x_j|z)}^{\xi_0 - 1}$, $j = 1, 2, \ldots, N$. These were Dirichlet distributions with hyperparameters which in this research was generated by using $\phi_0 > 0$ and $\xi_0 > 0$. The Dirichlet distribution [i.e., $\mathrm{Dir}(\alpha)$] is a family of continuous multivariate probability distributions parameterized by the vector α of positive reals, which can generate the multivariate generalization of the beta distribution, and conjugate prior of the categorical distribution and multinomial distribution in Bayesian statistics [1]. The Dirichlet distribution is the multinomial extension to the beta distribution for a binomial process, which can also be used in quantifying probabilities in predictive regression-based models [9].

We then let $\delta(n)$ be 1 when $n = 0$ and 1 otherwise, and then defined the sampled parameter uncertainty estimates by using $n_{(k,z_k)}^{z} := \sum_{i=1}^{n} \langle \delta(Z_i^{(k)} - z_k) \rangle_Q$. In this research, we also employed $n_{(j,x_j|z)}^{x} := \sum_{i=1}^{n} \delta\left(X_i^{(j)} - x_j\right) \left\langle \prod_{k=1}^{K} \delta(Z_i^{(k)} - z_k) \right\rangle_Q$. In these equations, $X_i^{(j)}$ was the state of the j-th observation node and

$Z_i^{(k)}$ was the state of the k-th hidden node. The variational posterior distribution of parameters [i.e., $\omega = \{a, b\}$] was then given by the equation $r(a_k|X^n) = \frac{\Gamma(n + T_k\phi_0)}{\prod_{z_{k=1}}^{T_k} \Gamma\left(n^z_{(k,z_k)} + \phi_0\right)} \prod_{z_{k=1}}^{T_k} a_{(k,z_k)}^{\overline{n}^x_{(k,z_k)}+\phi_0-1}$, where each of the sampled *S. damnosum s.l.* larval habitat endemic transmission oriented explanatory covariate coefficients were spatially quantitated employing the equation $^{-}n_{\downarrow}z^{\uparrow}x \,\square\, (:=)\sum_{\downarrow}(i=1)^{\uparrow} n \equiv \langle \prod_{\downarrow}(k=1)^{\uparrow} K \equiv [\delta(Z_{\downarrow}i^{\uparrow}((k)) - z_{\downarrow}k]\rangle_{\downarrow} Q.$

It then followed that $\overline{n}_z^x = \sum_{x_{j=1}}^{Y_j} \overline{n}^x_{(j,x_j|z)}$, for $j = 1, \ldots, N$, and $\overline{n}^z_{(k,z_k)} = \sum_{z_{-k}} \overline{n}_z^x$ which subsequently denoted the sum over $z_i (i \neq k)$.

In this research, we evaluated the statistical efficiency of the MCMC sequence in SASmacro WinBUGSio by using the following steps:

Step 1. Sample $S^{(m)}$ from $f(S|X, Z^{(m-1)})$;

Step 2. Sample $P^{(m)}$ from $f(P|X, S^{(m)}, Z^{(m-1)})$;

Step 3. Sample $Z^{(m)}$ from $f(Z|X, P^{(m)}, S^{(m)}, \Phi^{(m-1)})$;

where X was the field and remote-sampled *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related explanatory covariate coefficient and Z was randomly assigned a starting value $Z^{(1)}$ by using a uniform prior distribution.

Step 1 was performed by drawing from an inverse Wishart distribution. In Bayesian statistics, inverse Wishart distribution is used as the conjugate prior for the covariance matrix of a multivariate normal distribution [7]. The pdf of the inverse Wishart was: $\frac{[\psi|\square]^{\frac{m}{2}}|B|^{-\frac{m+p+1}{2}} e^{-\frac{trace(\psi B^{-1})}{2}}}{2^{\frac{mp}{2}} \tau_p\left(\frac{m}{2}\right)}$, where B and ψ were $p \times p$ positive definite matrices, and $\tau_p(\cdot)$ was the multivariate Gamma function. In

this research, the distribution of the inverse of a Wishart-distributed matrix was generated from $A \sim W(\sum, m)$ and $\sum$ which was of size $p \times p$ for the matrix constructed from the field and remote-sampled *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk-related-parameters. Under these circumstances, $B = A^{-1}$ had an inverse Wishart distribution $BW^{-1} \sim \sum \square^{-1}(, m)$. We calculated the pdf [i.e., $\tau_p(\cdot)$] which in this research represented the multivariate Gamma function of the sampled data. A pdf or density of a continuous random variable is a function that describes the relative likelihood for this random variable to occur at a given point [1].

The marginal and conditional distributions from the inverse Wishart-distributed matrix were then quantified using $\boldsymbol{A} \sim \boldsymbol{W}^{-1}(\psi, \boldsymbol{m})$. We then partitioned the matrices for determining if ψ was conformable with each other using $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$, $\psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}$, where $\mathbf{A}_{ij}$ and ψ_{ij} were $\boldsymbol{p}_i \times \boldsymbol{p}_j$ matrices. We then determined if:

(i) $\mathbf{A}_{11}$ was independent of $\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ and $\mathbf{A}_{22.1}$, when $\mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ was the Schur complement (i.e., a submatrix within a larger matrix) of $\mathbf{A}_{11}$ in $\mathbf{A}$;

(ii) $\mathbf{A}_{11\sim}W^{-1}(\psi_{11}, m - p_2)$;

(iii) $\mathbf{A}_{\downarrow}11^{\uparrow}(-1)\mathbf{A}_{\downarrow}12|\mathbf{A}_{\downarrow}(22 \cdot 1) \sim MN_{\downarrow}(p_{\downarrow}1xp_{\downarrow}2)(\psi_{\downarrow}11^{\uparrow}(-1)\psi_{\downarrow}12, \mathbf{A}_{\downarrow}(22 \cdot 1) \otimes \psi_{\downarrow}11^{\uparrow}(-1))$, where $MN_{p\times q}(\cdot,\cdot)$ was a matrix normal distribution generated from the regressed spatiotemporal-sampled *S. damnosum s.l.* larval habitat endemic transmission oriented explanatory covariate coefficients; and,

(iv) $\mathbf{A}_{22\cdot1} \sim W^{-1}(\psi_{11}, m)$.

We used the conjugate distribution to make inference about a covariance matrix $\sum\square$, whose prior had a $\boldsymbol{W}^{-1}(\psi_{11}, \boldsymbol{m})$ distribution. If the observations $\boldsymbol{X} = \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are independent p-variate Gaussian variables drawn from a distribution then the conditional distribution has a $\boldsymbol{W}^{-1}(\boldsymbol{A} + \psi, n + \boldsymbol{m})$ distribution, where $\mathbf{A} = XX^T$ is n times the sample covariance matrix [7]. Because the prior and posterior distributions are the same family, the inverse Wishart distribution was the conjugate to the multivariate Gaussian generated from the georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk-related data. This task was simplified by assuming that the riverine larval habitat endemic transmission-oriented predictive risk data analyses had covariance matrices with common eigenvectors. If covariance's differ among sources, the inverse Wishart draws often produce invalid results, especially for sources that are small components of the mixture [7]. As such, in this research, we developed covariance matrix S, which was decomposed into a vector of standard deviations, V, and a correlation matrix, R using: S=$diag$(V)$R diag$(V), where $diag$(V) was a matrix with diagonal elements V and zeros elsewhere. This decomposition permitted the independent sampling of V and R. We then simulated the standard deviations, V, from an inverse Gamma distribution: $p\left(V_{k,l}^2 | X, Z^{(m-1)}\right) \sim IG\left(\alpha + \frac{n_k}{2}, \beta + \frac{s_{k,l}^2 n_k}{2}\right)$, where n_k was the number of individual habitats assigned to cluster k by $Z^{(m-1)}$ (i.e., the *S. damnosum s.l.* larval habitat sample size), $s_{k,l}^2$ was the sample variance of element l in cluster k as assigned by $Z^{(m-1)}$, and α and β were constants, both set to the non-informative prior value of 1.

We then simulated the elements of the correlation matrices, R, from a hyperbolic-tangent transformed distribution using: $p\left(\tanh\left(R_{k,i,j}\right) | X, Z^{(m-1)}\right) \sim N\left(\tanh\left(\hat{R}_{k,i,j}\right), \frac{1}{n_k}\right)$, where $\hat{R}_{k,i,j}$ was the current estimate of the correlation between the i-th and j-th elements

$(i \neq j)$ in a cluster k given $Z^{(m-1)}$, and n_k was the same. After sampling values for both V and R, we reassembled the covariance matrix, S_K, for each habitat cluster, thus, completing Step 1.

Step 2 required drawing values for the elemental means, P. The field and remote-sampled *S. damnosum s.l.* riverine larval habitat data X had an approximate multivariate normal distribution as such Step 2 was performed by using the vector of mean concentrations for cluster k. The multivariate normal distribution using the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk explanatory covariates was then given by the sample means calculated from X, generated from the cluster assignments, $Z^{(m-1)}$ and the covariance S_k from Step 1. If cluster k was empty at $m-1$ (that was no individual sampled habitat parameters were assigned to k by $Z^{(m-1)}$), then the grand mean and covariance matrix of X were used as the sample mean and covariance matrix of k.

Step 3 involved drawing new cluster assignments using each individual sampled *S. damnosum s.l.* larval habitat estimators. To do so, it was necessary to calculate $\Pr(z_i = k)$ for every i, k combination (z_i was the i-th element of Z). Again, we assumed multivariate normal distributions where $Z^{(m)}$ was simulated from: $\Pr\left(z^{i,(m)} = k \middle| X, P^{(m)}, S^{(m)}, \Phi\right) = \dfrac{\phi_k f\left(X^i \middle| P_k^{(m)}, S_k^{(m)}, z^i = k\right)}{\sum_{K-1}^{K} \phi_{k'} f\left(X^i \middle| P_{k'}^{(m)}, S_{k'}^{(m)}, z^i = k'\right)}$, where the $f(.)$ terms on the right-hand side were multivariate normal likelihoods. Thus, the likelihood that the i-th element of X was present in the sampled larval habitat population k, was normalized by the sum of likelihoods for all potential sources influencing presence of immature *S. damnosum s.l.* at the Nabere study site. Finally, the sample $\Phi^{(m)}$ from $f(Z|X, P^{(m)}, S^{(m)}, Z^{(m)})$ was used to generate asymptotically efficient estimates in our seasonal multivariate *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk distribution model.

3. Results

Initially, the CLUSTER procedure in SAS hierarchically aggregated the *S. damnosum s.l.* endemic transmission-oriented predictive risk-based-observations. PROC CLUSTER then computed all Euclidean distances in the dataset based on the flexible-beta method. PROC CLUSTER then generated the number of clusters in the sampled *S. damnosum s.l.* larval habitat population. PROC CLUSTER also created an output dataset which was used by the TREE procedure in SAS to draw a diagram of the cluster hierarchy. In this research, to obtain the five-cluster solution, we first used PROC CLUSTER with the OUTTREE= option, and then employed the output data set as the input data set to the TREE procedure. Within PROC TREE, NCLUSTER specified the number of clusters based on the sampled georeferenced *S. damnosum s.l.* larval habitat data and the OUT= options to obtain the final solution and draw a tree diagram. Since we considered all the sampled georeferenced larval habitat explanatory covariates to be equally important, we used the STD option in PROC CLUSTER to standardize the variables to mean 0 and standard deviation 1. We removed the outliers before using PROC CLUSTER with the STD option. The STDIZE procedure provides additional methods for standardizing variables and imputing missing values (www.sas.edu).

In this research, the relationship between prevalence of each individual potential predictor variable sampled in the Nabere study site was investigated by single variable regression analysis in PROC MIXED. We used the regression line $(\mathbf{y}_i - \overline{\mathbf{y}}) = (\hat{\mathbf{y}}_i - \overline{\mathbf{y}}) + (\mathbf{y}_i - \hat{\mathbf{y}}_i)$ to generate a pseudo R^2 value where the first term was the total variation in the response y total (larval density count), and the second term was the variation in mean response based on the sampled parameters. The third term was the residual value in the model estimates. Squaring each of these terms and adding over all of the sampled *S. damnosum s.l.* riverine larval habitat observations generated the equation $\sum(y_i - \overline{y})^2 = \sum(\hat{y}_i - \overline{y})^2 + \sum(y_i - \hat{y}_i)^2$. This equation was then

written as SST = SSM + SSE, where SS was notation for sum of squares and *T*, *M*, and *E* were the notation for total, model, and error, respectively. By doing so, the square of the sample correlation was then equal to the ratio of the estimates, while the sum of squares was related to the total sum of squares: r^2 = SSM/SST. This formalized the interpretation of R^2 for explaining the fraction of variability in the sampled immature *S. damnosum s.l.* data explained by the regression model. The sample variance s_y^2 was then equal to $\sum \frac{(y_i - \bar{y})^2}{n-1}$, which in turn was equal to the SST/df, the total sum of squares divided by the total df. A regression equation was then constructed by using the mean square model (i.e., MSM) = $\sum \frac{(\hat{y}_i - \bar{y})^2}{l}$, which in this research was equal to the SSM/df. The corresponding MSE was $\sum \frac{(y_i - \hat{y}_i)^2}{n-2}$, we noticed was equal to SSE/df and the estimate of the variance about the regression line (i.e., σ^2). The MSE is an estimate of σ^2 for determining whether or not the null hypothesis is true [10].

For quantizing, the spatiotemporal-sampled explanatory predictor variables, (*p*) the *S. damnosum s.l.* larval habitat modeled the DFM, which we noted was equal to *p* and the error degrees of freedom (DFE). This product was also equal to $(n - p - 1)$, and the total degrees of freedom (DFT), which was subsequently equal to $(n - 1)$, the sum of DFM and DFE. We also noted that all the explanatory and response variables were numeric. The relationship between the mean of the response variable (i.e., total larval density count) and the level of the sampled explanatory covariates in the regression equation were assumed to be approximately linear (i.e., straight line). The corresponding table generated classified each the field and remote-sampled *S. damnosum s.l.* larval habitat parameter estimator in SAS see Table 2.

Table 2. The *S. damnosum s.l.* larval habitat regression-based model parameter estimates

Source	Sum of Squares	Mean Square	
Model	p MSM/MSE	$\sum(\hat{\mathbf{y}}_i - \bar{\mathbf{y}})^2$	SSM/DFM
Error	$n - p - 1$	$\sum(y_i - \hat{\mathbf{y}}_i)^2$	SSE/DFE
Total	$n - 1$	$\sum(y_i - \bar{\mathbf{y}})^2$	SST/DFT

In the multiple regression analyses, the test statistic MSM/MSE had an $F(p, n - p - 1)$ distribution. In this research, the null hypothesis was $\beta_1 = \beta_2 = \ldots = \beta_p = 0$, and the alternative hypothesis was at least one of the sampled *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk parameters $\beta_j \neq 0$, $j = 1, 2, \ldots, p$. The F test did not indicate which of the parameters $\beta_j \neq 0$ nor which was not equal to zero only that at least one of them was linearly related to the response variable. The ratio SSM/SST $= R^2$ (i.e., squared multiple correlation coefficient) was thereafter the proportion of the variation in the response variable that was explained by the immature sampled habitat data. The square root of R^2 (i.e., the multiple correlation coefficient) was then the correlation between the sampled *S. damnosum s.l.* larval habitat observations (i.e., y_i) and the fitted values (i.e., $\hat{\mathbf{y}}_i$). Additionally, from the sampling distribution generated from the sampled t parameters, the probability of obtaining an F was large or larger than the one was calculated. In this research, there were only two means to compare, the t-test and the F-test, which were equivalent. The relation between ANOVA and t was then given by $F = t^2$. Thereafter, significant differences by ANOVA were noted for mean numbers of the *S. damnosum s.l.* captured throughout the sampling frame ($F = 44.7$, $DF = 1$).

A Poisson regression analyses was then constructed in PROC MIXED to determine the relationship between the sampled *S. damnosum s.l.* larval habitat count data and the sampled habitat characteristics. The

Poisson models were built by using the field and remote-sampled *S. damnosum s.l.* larval habitat multivariate endemic transmission-oriented predictive risk data. A negative binomial regression had to be employed, however, as an examination of the data indicated that overdispersion was a significant problem in the Poisson model. The Poisson distribution is a special case of the negative binomial distribution where the mean approximates the standard deviation [8]. We assumed that the log of the mean μ was a linear function of independent variables, $\log(\mu) = \text{intercept} + b1^{*}X1 + b2^{*}X2 + \ldots + b3^{*}Xm$ in the model, which implied that μ was the exponential function of independent variables when $\mu = \exp(\text{intercept} + b1^{*}X1 + b2^{*}X2 + \ldots + b3^{*}Xm)$. Therefore, instead of assuming that the distribution of the sampled *S. damnosum s.l.* larval habitats parameter estimates (i.e., Y) was Poisson, we were able to assume that Y had a negative binomial distribution. We relaxed the assumption about equality of mean and variance (i.e., Poisson distribution property), since the variance of negative binomial was equal to $\mu + k\mu^2$, where $k >= 0$ was a dispersion parameter. The ML method was then used to estimate k, as well as the sampled larval habitat parameters of the regression model for $\log(\mu)$. For the negative binomial distribution, the variance was equal to the mean $+ k\,\text{mean}^2$ (i.e., $k >= 0$) as the negative binomial distribution reduced to Poisson when k was 0.

In the regression analyses, the null hypothesis was: $H_0 : k = 0$ and the alternative hypothesis was: $H_a : k > 0$. We recorded the log-likelihood (i.e., LL) for the models. We employed the likelihood ratio (LR) test to compute the LR statistic using $-2(\text{LL})$ (Poisson) and the LL (i.e., negative binomial). The asymptotic distribution of the LR statistic had a probability mass of one half at zero and one half - Chi-square distribution with 1 df. To test the null hypothesis at the significance level α, we used the critical value of Chi-square distribution corresponding to significance level 2α, whereby there was a rejection of H_0 if LR statistic $> \chi^2_{(1-2\alpha,1\,\text{df})}$. We generated the log of the mean, μ using a linear function of independent variables whereby $\log(\mu) = \text{intercept} + b1^{*}X1 + b2^{*}X2$

$+ \ldots + b3^{*} Xm$, in the *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk model implied that μ was the exponential function of the independent variables when $\mu = \exp(\text{intercept} + b1^{*} X1 + b2^{*} X2 + \ldots + b3^{*} Xm)$. By so doing, distance from the capture point was found to be significantly associated to the sampled *S. damnosum s.l.* riverine larval habitats at the Nabere study site.

All residual estimates from the Bayesian model were then evaluated in a spatial error (SE) model. An autoregressive model was employed that used a sampled habitat variable, *Y*, as a function of nearby sampled habitat *Y* values [i.e., an autoregressive response (AR) or spatial linear (SL) specification] and/or the residuals of *Y* as a function of nearby *Y* residuals [i.e., an AR or SE specification]. Distance between sampled habitats was then defined in terms of an *n*-by-*n* geographic weights matrix, **C**, whose c_{ij} values were 1 if the sampled riverine larval habitat locations i and j were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum, with the row sums given by **C1**, converted this matrix to matrix **W**. The *n*-by-1 vector $x = [x_1 \cdots x_n]^T$ then contained measurements of a quantitative variable for n spatial units and *n*-by-*n* spatial weighting matrix **W**. The formulation for the Moran's index of spatial autocorrelation used in this research was: $I(x) = \dfrac{n\sum_{(2)} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{(2)} w_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2}$, where $\sum_{(2)} \sum_{i=1}^{n} \sum_{j=1}^{n}$ with $i \neq j$. The values w_{ij} were spatial weights stored in the symmetrical matrix **W** [i.e., $(w_{ij} = w_{ji})$] that had a null diagonal $(w_{ii} = 0)$. In this research, the matrix was generalized to an asymmetrical matrix **W**. Matrix **W** can be generalized by a non-symmetric matrix W^{*} by using $W = (W^{*} + W^{*T})/2$ [11]. Moran's I was the rewritten using matrix notation:

$$I(x) = \frac{n}{1^{t} W_1} \frac{x^T HHWHHx}{x^T HHx} = \frac{n}{1^T W_1} \frac{x^T HWHx}{x^T Hx}, \quad \text{where} \quad H = (I - 11^T/n)$$

was an orthogonal projector verifying that $H = H^2$ (i.e., H was

independent). Features of matrix $\mathbf{W}$ for analyzing the endemic transmission oriented explanatory sampled covariate coefficients of the *S. damnosum s.l.* riverine larval habitats included that it was a stochastic matrix which expressed each observed value *yi* as a function of the average of habitat location *i*'s nearby habitat larval counts, while allowing for a single spatial autoregressive parameter ρ to have a maximum value of 1.

An SAR model specification was then used to describe the *S. damnosum s.l.* riverine larval habitat model autoregressive variance uncertainty estimates. A spatial filter (SF) model specification was also used to describe both Gaussian and Poisson random variables. The resulting SAR model specification took on the following form: $\mathbf{Y} = \mu(1-\rho)\mathbf{1} + \rho\mathbf{WY} + \varepsilon$, where μ was the scalar conditional mean of *Y*, and ε was an *n*-by-1 error vector, whose elements were statistically independent normally random variates.

The spatial covariance matrix for Equation (2.1) then incorporated the sampled riverine larval habitat endemic transmission oriented explanatory covariates in $E[(\mathbf{Y}-\mu\mathbf{I})'(\mathbf{Y}-\mu\mathbf{I})] = \Sigma = [(\mathbf{I}-\rho\mathbf{W}')(\mathbf{I}-\rho\mathbf{W})]^{-1}\sigma^2$, where $E(\bullet)$ denoted the calculus of expectations, $\mathbf{I}$, which was the *n*-by-*n* identity matrix denoting the matrix transpose operation where σ^2 was the error variance. However, when a mixture of PSA and NSA is present in a vector arthropod-related larval habitat model, a more explicit representation of both effects leads to a more accurate interpretation of empirical results [2]. Alternately, the excluded values may be set to zero, although if this is done then the mean and variance must be adjusted.

In this research, two different spatial autoregressive parameters appeared in the spatial covariance matrix riverine larval habitat model specification which for our SAR model specification became: $\Sigma = [(\mathbf{I}- < \rho >_{diag} \mathbf{W}')(\mathbf{I}- < \rho >_{diag} \mathbf{W})]^{-1}\sigma^2$, where the diagonal matrix of autoregressive parameters, $< \rho >_{diag}$, contained two sampled parameters: ρ_+ for those specific habitats. The habitat pairs displayed positive spatial dependency (i.e., ρ, for those habitat pairs displaying

negative spatial dependency). For example, by letting $\sigma^2 = 1$ and employing a 2-by-2 regular square tessellation rendered

$$\Sigma=\left[\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}-\begin{pmatrix}\rho_+&0&0&0\\0&\rho_+&0&0\\0&0&\rho_-&0\\0&0&0&\rho_-\end{pmatrix}\begin{pmatrix}0&\frac{1}{2}&\frac{1}{2}&0\\\frac{1}{2}&0&0&\frac{1}{2}\\\frac{1}{2}&0&0&\frac{1}{2}\\0&\frac{1}{2}&\frac{1}{2}&0\end{pmatrix}\right]^2.$$

For the vector $\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}$, this matrix enabled positing a positive relationship between the sampled *S. damnosum s.l.* habitats by endemic transmission oriented covariates, y_1 and y_2. Additionally a negative relationship between the covariates, y_3 and y_4, and, no relationship between covariates, y_1 and y_3 and between y_2 and y_4 were noted. This covariance specification yielded: $\mathbf{Y} = \mu(\mathbf{I} - \rho_+ < \mathbf{I}_+ >_{diag} - \rho_- < \mathbf{I}_- >_{diag})\mathbf{1} + (\rho_+ < \mathbf{I}_+ >_{diag} + \rho_- < \mathbf{I}_- >_{diag})\mathbf{WY} + \varepsilon$, where $\mathbf{I}_+$ was a binary 0-1 indicator variable. This equation also denoted those *S. damnosum s.l.* riverine larval habitat covariates displaying positive spatial dependency where $\mathbf{I}_-$ was a binary 0-1 indicator variable denoting those sampled habitats displaying negative spatial dependency using $\mathbf{I}_+ + \mathbf{I}_- = 1$. Expressing the preceding 2-by-2 example in terms of Equation (2.3) yielded:

$$\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}=\mu\left[\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}-\rho_+\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}-\rho_-\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\right]$$

$$\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \left[\rho_+ \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix} + \rho_- \begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\right]$$

$$\begin{pmatrix}0&\frac{1}{2}&\frac{1}{2}&0\\\frac{1}{2}&0&0&\frac{1}{2}\\\frac{1}{2}&0&0&\frac{1}{2}\\0&\frac{1}{2}&\frac{1}{2}&0\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix} + \begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\end{pmatrix}.$$

If either $\rho_+ = 0$ (and hence $\mathbf{I}_+ = \mathbf{0}$ and $\mathbf{I}_- = \mathbf{I}$) or $\rho_- = 0$ (and hence $\mathbf{I}_- = \mathbf{0}$ and $\mathbf{I}_+ = \mathbf{I}$), appeared in the model then Equation (2.3) reduced to Equation (2.1). This indicator variables classification was made in accordance with the quadrants of the corresponding Moran scatterplot generated by using the sampled *S. damnosum s.l.* larval habitat endemic transmission oriented covariate coefficients sampled in the riverine epidemiological study site. In our model PSA and NSA processes counterbalanced each other in a mixture such that the sum of the two spatial autocorrelation parameters-$(\rho_+ + \rho_-)$ were close to 0. Further, the Jacobian estimation was implemented by utilizing the differenced indicator riverine larval habitat variables in $\mathbf{I} - \gamma\mathbf{I}$ for estimating ρ and γ with ML techniques, and setting $\hat{\rho}_- = -\hat{\gamma}\hat{\rho}_+$. The Jacobian then generalized the gradient of a scalar valued function of multiple variables by generalizing the derivative of a scalar-valued function of a scalar. A more complex *S. damnosum s.l.* riverine larval habitat specification was then posited by generalizing these binary indicator variables. We used $F : R^n \to R^m$ as a function from Euclidean n-space to Euclidean m-space, which was simulated in SAS using the distance measurements between the sampled *S. damnosum s.l.*riverine larval habitats. This function was given by m larval habitat covariate coefficients

(i.e., component functions) as described by $y_1(x_1, x_n)$, $y_m(x_1, x_n)$. The partial derivatives of all these functions were then organized in an m-by-n matrix whereby, the Jacobian matrix J of F was as follows:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

This matrix was denoted by $J_F(x_1, a_n)$ and $\frac{\partial(y_1, \ldots, y_m)}{\partial(x_1, \ldots, x_n)}$. The i-th row $(i = 1, m)$ of this matrix was the gradient of the i-th component function $y_i : (\nabla y_i)$. In these analyses, **p** was a sampled riverine habitat covariate in R^n and F (i.e., sampled larval count) differentiable at **p**. A derivative was then given by $J_F(p)$. The model described by $J_F(p)$ was the best linear approximation of F near the larval habitat point **p** in the sense that: $F(x) = F(p) + J_F(p)(x - p) + o(\|x - p\|)$. The spatial structuring was then achieved by constructing a linear combination of a subset of the eigenvectors of a modified geographic weights matrix using $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ that appeared in the numerator of the MC spatial autocorrelation which was indexed with a product moment correlation coefficient.

A subset of eigenvectors was then selected with a stepwise regression procedure. Because $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}'/n) = \mathbf{E}\Lambda\mathbf{E}'$, where **E** is an n-by-n matrix of eigenvectors and Λ is an n-by-n diagonal matrix of the corresponding eigenvalues [11] in this research, resulting *S. damnosum s.l.* larval habitat model specification was given by: $\mathbf{Y} = \mu\mathbf{1} + \mathbf{E}_k\beta + \varepsilon$, where μ was the scalar mean of Y, E_k was an n-by-k matrix containing the subset of $k <<_n$ eigenvectors selected with a stepwise regression technique where β was a k-by-1 vector of regression coefficients.

A number of the eigenvectors were then extracted from $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\,\mathbf{C}\,(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$, which were affiliated with geographic patterns of the sampled *S. damnosum s.l.* habitat endemic transmission oriented explanatory covariates at the study site, portraying a negligible degree of spatial autocorrelation. Consequently, only k of the n eigenvectors was of interest for generating a candidate set for a stepwise regression procedure. Candidate eigenvector represents a level of spatial autocorrelation, which can account for the redundant information in orthogonal riverine larval habitat map patterns [2].

Of note, the 2-by-2 square tessellation rendered a repeated eigenvalue in our model. To identify spatial clusters of *S. damnosum s.l.* riverine larval habitats then a Thiessen polygon surface partitioning was generated to construct geographic neighbour matrices which also were employed in the spatial autocorrelation analysis. Entries in matrix were 1, if two sampled riverine larval habitats shared a common Thiessen polygon boundary and 0, otherwise. Next, the linkage structure for each surface was edited to remove unlikely geographic neighbours to identify pairs of sampled larval habitats sharing a common Thiessen polygon boundary. Attention was restricted to those map patterns associated with at least a minimum level of spatial autocorrelation, which, for implementation purposes, was defined by $|MC_j / MC_{\max}| > 0.25$, where MC_j denoted the j-th value and $MC_{\max}$, the maximum value of MC. This threshold value allowed two candidate sets of eigenvectors to be considered for substantial PSA and substantial NSA, respectively. These statistics indicated that the detected NSA may be considered to be statistically significant in a regressed dataset of the *S. damnosum s.l.* riverine larval habitat endemic transmission oriented explanatory covariate coefficients, based upon a randomization perspective. Of note, was that the ratio of the PRESS (i.e., predicted error sum of squares) statistic to the sum of squared errors from the MC scatterplot trend line which in this research was 1.27. Fortunately, was well within two standard deviations of the average standard prediction error value (roughly 1.18) for a sampled *S. damnosum s.l.* larval habitat at the study site. Because larval counts were being analyzed, a Poisson spatial filter model specification was employed in this research, thereafter.

The model specification was then written as follows: $LN(\mu) = \alpha\mathbf{1} + \mathbf{E}_k\beta$, $\sigma_i^{\swarrow} = \mu_i(1 - \eta\mu_i)$, where μ_i was the expected mean larval count for a georeferenced *S. damnosum s.l.*riverine larval habitat location i, where μ was an n-by-1 vector of the expected larval counts. In this research, LN denoted the natural logarithm (i.e., the generalized linear model link function), α was an intercept term and η was the negative binomial dispersion parameter. This log-linear equation had no error term; rather, estimation was executed assuming a negative binomial random variable.

The eigenfunctions of a spatial weighted *S. damnosum s.l.* riverine larval habitat matrix was then generated. The upper and lower bounds for a spatial matrix was then derived by using Moran's indices (I) given by $\lambda_{\max}(n/1^T W1)$ and $\lambda_{\min}(n/1^T W1)$, where $\lambda_{\max}$ and $\lambda_{\min}$ were the extreme eigenvalues of $\Omega = HWH$. Hence, in this research, the eigenvectors of Ω were vectors with unit norm maximizing Moran's I. The eigenvalues of this matrix were equal to Moran's I coefficients of spatial autocorrelation post-multiplied by a constant. Eigenvectors associated with high positive (or negative) eigenvalues have high positive (or negative) autocorrelation [11].

We noticed that eigenvectors associated with eigenvalues with extremely small absolute values corresponded to low spatial autocorrelation which were not suitable for defining spatial structures in the *S. damnosum s.l.* riverine larval habitat model. The diagonalization of the spatial weighting matrix rendered from the field and remote-sampled endemic transmission oriented explanatory covariate coefficients instead consisted of finding the normalized vectors u_i stored as columns in the matrix $U = [u_1 \cdots u_n]$, for satisfying: $\Omega = HWH = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$, where $\Lambda = diag(\lambda_1 \cdots \lambda_n)$, $u_i^T u_i = \|u_i\|^2 = 1$ and $u_i^T u_j = 0$ when $i \neq j$. Note that double centering of Ω implied that the eigenvectors u_i generated from the ecological sampled regressed larval habitat covariates

were centered whereby at least one eigenvalue was equal to zero. Introducing these eigenvectors in the original formulation of Moran's index lead to

$$I(x) = \frac{n}{1^T W 1} \frac{x^T HWHx}{x^T Hx} = \frac{n}{1^T W 1} \frac{x^T U \Lambda U^T x}{x^T Hx} = \frac{n}{1^T W 1} \frac{\sum_{i=1}^{n} \lambda_i x^T u_i u_i^T x}{x^T Hx}. \tag{3.1}$$

Considering the centered vector $z = Hx$ and using the properties of idempotence of H, Equation (3.1) was equivalent to

$$I(x) = \frac{n}{1^T W 1} \frac{\sum_{i=1}^{n} \lambda_i z^T u_t u_i^T z}{z^T z} = \frac{n}{1^T W 1} \frac{\sum_{i=1}^{n} \lambda_i \left\| u_i^T z \right\|^2}{\|z\|^2}. \tag{3.2}$$

We then transformed the autocorrelation indicators to correlation coefficient as the eigenvectors u_i and the vector z were then centered and the *S. damnosum s.l.* predictive equation was rewritten as:

$$I(x) = \frac{n}{1^T W 1} \frac{\sum_{i=1}^{n} \lambda_i cor^2(u_i, z) \operatorname{var}(z) n}{\operatorname{var}(z) n} = \frac{n}{1^T W 1} \sum_{i=1}^{n} \lambda_i cor^2(u_i, z). \tag{3.3}$$

In this research, r was the number of null eigenvalues of $\Omega (r \geq 1)$. These eigenvalues and corresponding eigenvectors were removed from Λ and U, respectively. The residual forecasts were then strictly equivalent to $I(x) = \frac{n}{1^T W 1} \sum_{i=1}^{n-r} \lambda_i cor^2(u_i, z)$. Moreover, it was demonstrated that Moran's index for a given eigenvector u_i was equal to $I(u_i) = (n/1^T W 1)\lambda_i$, so the equation was rewritten: $I(x) = \sum_{i=1}^{n-r} I(u_i) cor^2(u_i, z)$. The term $cor^2(u_i, z)$ represented the part of the variance of z that was explained by u_i in the *S. damnosum s.l.* riverine larval habitat model [i.e., $z = \beta_i u_i + e_i$]. This quantity was equal to $\beta_i^2 / n \operatorname{var}(z)$. By definition, the eigenvectors

u_i were orthogonal and therefore in this research regression coefficients of the linear models $z = \beta_i u_i + e_i$ were those of the multiple regression model $z = U\beta + \varepsilon = \beta_i u_i + \ldots + \beta_{n-r} u_{n-r} + \varepsilon$.

Next, the distribution of the error residuals in the *S. damnosum s.l.* larval habitat autocovariance matrix revealed that the maximum value of I was obtained by all of the variation of z, as explained by the eigenvector u_1, which coincidentally corresponded to the highest eigenvalue λ_1 in the spatial autocorrelation error matrix. We found that $cor^2(u_i, z) = 1$ and $cor^2(u_i, z) = 0$ for $i \neq 1$. We then noted that the maximum value of I was deduced for Equation (3.3), which was equal to $I_{MAX} = \lambda_1(n/1^T W1)$. This minimum value in the error matrix was thereafter quantitated based on all the variation of z, which was explained by the eigenvector u_{n-r} corresponding to the lowest eigenvalue λ_{n-r} generated in the riverine larval habitat model. This minimum value was equal to $I_{\min} = \lambda_{n-r}(n/1^T W1)$. If the ecological sampled predictor variable was not spatialized, the part of the variance explained by each eigenvector was then equal to $cor^2(u_i, z) = 1/n - 1$. Because the field and remote-sampled *S. damnosum s.l.* riverine larval habitat variables in z were randomly permuted, it was assumed that we would obtain this result. In this research, the set of $n!$ random permutations revealed that $E_R(I) = \frac{n}{1^T W1(n-1)} \sum_{i=1}^{n} \lambda_i = \frac{n}{1^T W1(n-1)} trace(\Omega)$. It was thereafter easily demonstrated that $trace(\Omega) = -\frac{1^T W1}{n}$; thus, it followed that $E_R(I) = -\frac{1}{n-1}$.

The sampled georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission oriented explanatory predictor covariates were then imput into an eigenfunction decomposition algorithm to quantitate any autocorrelation error coefficients in the linear residual variance estimates. Results indicated that negligible PSA was detected for the

sampled *S. damnosum s.l.* larval habitat data. Eigenvectors were then extracted from the matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$, using the ecological-sampled predictor variables. In this research denoting the autoregressive parameter captured the latent spatiotemporal autocorrelation in the seasonal multivariate *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk model. This quantification involved ρ, a conditional autoregressive covariance specification and the matrix $(\mathbf{I}^{-}\rho\mathbf{C})$, where $\mathbf{I}$ was an n-by-n identity matrix. Thereafter, the residual autocorrelation error components were calculated as the matrix $\mathbf{C}$ raised to the power 1. Further, since adjacent sampled larval habitat data were involved in the autoregressive function, we used a first-order specification, with the autoregressive term being $\mathbf{CY}$. An important matrix was then derived from $\mathbf{C1}$, which was the vector of the number of sampled *S. damnosum s.l.* larval habitat neighbours in the study site. In this research, the inverse of the elements of $\mathbf{C1}$ were inserted into the diagonal of a diagonal matrix (i.e., $\mathbf{D}^{-1}$) rendering matrix $\mathbf{W} = \mathbf{D}^{-1}\mathbf{C}$, which became a stochastic matrix (i.e., each of its row sums equaled 1). One appealing feature of this matrix was that the autoregressive term became $\mathbf{WY}$, which generated averages, rather than sums, of the neighbouring sampled larval habitat parameter estimate values. Because a covariance matrix for a robust vector larval habitat distribution model must be symmetrical [2], we employed a matrix $\mathbf{W}$ specification with a conditional autoregressive model by making the individual-sampled larval habitat variance nonconstant using $(\mathbf{I} - \rho\mathbf{D}^{-1}\mathbf{C})\mathbf{D}^{-1} = (\mathbf{D}^{-1} - \rho\mathbf{D}^{-1}\mathbf{C}\mathbf{D}^{-1})$. An appealing feature of this version for our *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk distribution model was that it restricted values of the autoregressive parameter to the more intuitively interpretable range of $0 \le \hat{\rho} \le 1$. The larval habitat model then furnished an alternative specification, which was also written in terms of matrix $\mathbf{W}$. The spatial covariance was then a function of the matrix $(\mathbf{I} - \rho\mathbf{C}\mathbf{D}^{-1})(\mathbf{I} - \rho\mathbf{D}^{-1}\mathbf{C}) = (\mathbf{I} - \rho\mathbf{W}^{T})(\mathbf{I} - \rho\mathbf{W})$, where T denoted the matrix transpose. The resulting matrix was symmetric and was

considered a second-order specification as it included the product of two spatial structure matrices (i.e., $\mathbf{W}^T\mathbf{W}$). This matrix restricted values of the autoregressive parameter to the more intuitively interpretable range of $0 \leq \hat{\rho} \leq 1$.

Estimation results for these models appear in Table 3. PSA and NSA spatial filter component pseudo-R^2 values were reported. These values did not exactly sum for the complete spatial filter; however, they were very close to their corresponding totals, suggesting that any induced multicollinearity was quite small.

Table 3. Global spatial analyses of the sampled georeferenced *S. damnosum s.l.* larval habitat count data in the Nabere study site

Study Site	*n*	Transformation	MC	GR
Nab ere	152	LN (count +1)	0.067	0.891

A GLM was then extended to account for latent non-spatial correlation effects, which allowed inferences to be drawn for a much wider range of geographic sampling configurations generated from the sampled georeferenced *S. damnosum s.l.* larval habitat endemic transmission oriented explanatory covariate coefficients than those utilized by employing a GLMM. The GLMM included a random effect, which was specified as a random intercept that was assumed to be normally distributed with a mean of zero, a constant variance, and zero spatial autocorrelation. This varying intercept term compensated for the non-constant mean associated with the negative binomial model generalized specification. The spatial structuring of random effects was then implemented with a conditional autoregressive model and was achieved in this research with a spatial filter. The spatial autocorrelation components revealed 11% redundant information in the ecologically sampled datasets GLMM estimation results which appear in Table 4.

Table 4. Poisson spatial filtering model results for the sampled *S. damnosum s.l.* larval count data in the Nabere study site

Spatial Statistics	Nabere
SF: # of eigenvectors	5
SF: MC	0.634
SF: GR	0.462
SF pseudo- R^2	0.163
Positive SA SF: # of eigenvectors	3
Positive SA SF: MC	0.564

SF denotes spatial filter.

SA denotes spatial autocorrelation.

Table 5 lists the improvements of fit in the adjusted and unadjusted models for all model specifications and random error in the spatial analyses. The unadjusted model compared the univariate model to a model containing only the intercept term. Interactions were examined, and significant interactions were included. Improvement of fit was also calculated for the first-order interaction models to determine whether including significant interactions improved fit compared to the full main effects model. Convergence problems prevented obtaining results of a saturated model to determine whether the presented model fit as well as the saturated model.

Table 5. Comparison of improvement of fit measured by likelihood ratio between unadjusted and adjusted effects models

Unadjusted effects				Adjusted effects		
		Improvement			Improvement	
Variable	Deviance	χ^2	df	Deviance	χ^2	df
Intercept	983.1241					
AGVEG	983.5936	12.1098	1	885.169	3.7497	1
HGVEG	982.6438	14.1147	1	896.257	19.397	1
DDVEG	986.3168	8.00961	1	890.007	8.6375	1
DISBTHAB	986.5872	9.96053	1	901.328	20.632	1
Full main effects						
1st degree interactions				844.541	38.9926	5

In this research, we generated the inverse Wishart distribution, which was a probability distribution defined by the positive-definite matrix generated from the *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related-explanatory covariates. In our Bayesian model, we employed the inverse Wishart distribution to generate the conjugate prior for the covariance matrix of a multivariate normal distribution. In Bayesian probability theory, if the posterior distributions $p(\theta|x)$ are in the same family as the prior probability distribution $p(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood [7]. In this research, the pdf of the inverse Wishart was: $\frac{|\Psi|^{m/2}|\mathbf{B}|^{-(m+p+1)/2}e^{-\text{trace}(\Psi\mathbf{B}^{-1})/2}}{2^{mp/2}\Gamma_p(m/2)}$, where $\mathbf{B}$ and Ψ were $p \times p$ positive definite matrices, and $\Gamma_p(\cdot)$ was the multivariate Gamma function. If B follows an inverse Wishart distribution, denoted as $\mathbf{B} \sim \mathbf{W}^{-1}(\Psi, m)$, its inverse $\mathbf{B}^{-1}$ has a Wishart distribution $\mathbf{W}(\Psi^{-1}, m)$.

In this research, the multivariate Gamma function, $\Gamma_p(\cdot)$, was a generalization of the Gamma function in the *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk model. The Gamma function is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers and, as such, n is a positive integer: $\Gamma(n) = (n-1)!$ [1]. The Gamma function appears commonly in the pdf of the Wishart and inverse Wishart distributions [7].

In this research, the distribution generated from the sampled *S. damnosum s.l.* larval habitats data had an inverse Wishart distribution $\mathbf{A} \sim \mathbf{W}^{-1}(\Psi, m)$. We then successfully partitioned the matrices $\mathbf{A}$ and Ψ using $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$, $\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}$, where $\mathbf{A}_{ij}$ and Ψ_{ij} were $p_i a \times p_j$ matrices. By doing so, we generated the conjugate distribution of the sampled georeferenced endemic transmission oriented explanatory

covariates by using a covariance matrix $\sum$, whose prior $p(\sum)$ had a $\mathbf{W}^{-1}(\Psi, m)$ distribution. The sampled *S. damnosum s.l.* larval habitats observations were independent p-variate Gaussian variables which were drawn from a $\mathbf{N}(\mathbf{0}, \sum)$ distribution. The conditional distribution [i.e., $p(\sum|X)$] had a $\mathbf{W}^{-1}(\mathbf{A}+\Psi, n+m)$ distribution. Thereafter, $\mathbf{A} = \mathbf{X}\mathbf{X}^T$ was used to generate the sample covariance matrix. In this research, the inverse Wishart distribution was conjugate to the multivariate Gaussian. Due to its conjugacy to the multivariate Gaussian, it was possible to "integrate out" the Gaussian-based *S. damnosum s.l.* larval habitat parameters [i.e., $\sum$] from the other predictor variables. As such we used:

$$P(\mathbf{X}|\Psi, m) = \int P(\mathbf{X}|\,\Sigma)P(\Sigma|\Psi, m)d\,\Sigma = \frac{|\Psi|^{\frac{m}{2}}\Gamma_p\left(\frac{m+n}{2}\right)}{\pi^{\frac{np}{2}}|\Psi + \mathbf{A}|^{\frac{m+n}{2}}\Gamma_p\left(\frac{m}{2}\right)}.$$ In this research, the variance matrix $\sum$ for the sampled riverine larval habitat endemic transmission oriented explanatory covariates was Ψ (i.e., the priori) and as such $\mathbf{A}$ was directly obtained from the coefficient indicator values. The mean in the *S. damnosum s.l.* larval habitat model was then $\mathbf{E}(\mathbf{B}) = \frac{\Psi}{m-p-1}$. The variance of each element of $\mathbf{B}$ was then $\mathrm{var}(b_{ij}) = \frac{(m-p+1)\psi_{ij}^2 + (m-p-1)\psi_{ii}\psi_{jj}}{(m-p)(m-p-1)^2(m-p-3)}$. The variance of the diagonal used in the larval habitat distribution model was also rendered by using the same formula as above with $i = j$, which further simplified the model to: $\mathrm{var}(b_{ii}) = \frac{2\psi_{ii}^2}{(m-p-1)^2(m-p-3)}$.

We then performed a decomposition of a square matrix (i.e., A) into eigenvalues and eigenvectors (i.e., eigendecomposition). We defined a right eigenvector as a column vector $\mathbf{X}_R$ satisfying $A\mathbf{X}_R = \lambda_R\mathbf{X}_R$, where $\boldsymbol{A}$ was a matrix so $(\mathbf{A} - \lambda_R\,|)\mathbf{X}_R = \mathbf{0}$, which meant the right eigenvalues had zero determinant. Similarly, we defined a left

eigenvector as a row vector $\mathbf{X}_L$ satisfying $\mathbf{X}_L\mathbf{A} = \lambda_L\mathbf{X}_L$. Taking the transpose of each side rendered $(\mathbf{X}_L\mathbf{A})^T = \lambda_L{\mathbf{X}_L}^T$, which in this research was rewritten as $\mathbf{A}^T{\mathbf{X}_L}^T = \lambda_L{\mathbf{X}_L}^T$. We then rearranged this equation once again to obtain $\left(\mathbf{A}^T - \lambda_L\,|\right){\mathbf{X}_L}^T = \mathbf{0}$, which generated $\det\left(\mathbf{A}^T - \lambda_L\,|\right) = \mathbf{0}$. The equation, in turn, generated $0 = \det\left(\mathbf{A}^T - \lambda_L\,|\right) = \det\left(\mathbf{A}^T - \lambda_L\,|^T\right)$, $\det(\mathbf{A} - \lambda_L\,|)^T \det(\mathbf{A} - \lambda_L\,|)$, where the last step was from the identity was $\det(\mathbf{A}) = \det\left(\mathbf{A}^T\right)$. We equated these equations to 0 for *A* and *X*, which required that $\lambda_R = \lambda_L \equiv \lambda$ (see [7]). We then let $\mathbf{X}_R$ be a matrix formed by the columns of the right eigenvectors $\mathbf{X}_L$, which was actually a matrix formed by the rows of the left eigenvectors. We then let $D \equiv \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$, and as such $\mathbf{A}\mathbf{X}_R = \mathbf{X}_R D$ and $\mathbf{X}_L\mathbf{A} = D\mathbf{X}_L$ and $\mathbf{X}_L\mathbf{A}\mathbf{X}_R = \mathbf{X}_L\mathbf{X}_R D$ while, $\mathbf{X}_L\mathbf{A}\mathbf{X}_R = D\mathbf{X}_L\mathbf{X}_R$, so $\mathbf{X}_L\mathbf{X}_R D = D\mathbf{X}_L\mathbf{X}_R$. But this equation was of the form CD=DC, where **D** was a diagonal matrix, so, therefore, $\mathbf{C} \equiv \mathbf{X}_L\mathbf{X}_R$ was also diagonal. If *A* is a symmetric matrix, then the left and right eigenvectors are simply each other's transpose, and if *A* is a self-adjoint matrix (i.e., Hermitian), then the left and right eigenvectors are adjoint matrices [9]. In predictive autoregressive vector habitat modeling, a Hermitian matrix is a square matrix with complex entries that is equal to its own conjugate transpose – that is, the element in the *i*-th row and *j*-th column is equal to the complex conjugate of the element in the *j*-th row and *i*-th column, for all indices i and j: $a_{i,\,j} = \overline{a_{j,i}}$ [8]. Using the matrix with eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and corresponding eigenvalues λ_1, λ_2, and λ_3, an arbitrary vector $\mathbf{y}$ was then written as $\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3$. In this research, matrix *A* was generated $\mathbf{A}\mathbf{y} = b_1\mathbf{A}\mathbf{x}_1 + b_2\mathbf{A}\mathbf{x}_2 + b_3\mathbf{A}\mathbf{x}_3 = \lambda_1\left(b_1\mathbf{x}_1 + \frac{\lambda_2}{\lambda_1}b_2\mathbf{x}_2 + \frac{\lambda_3}{\lambda_1}b_3\mathbf{x}_3\right)$; so $\mathbf{A}^n\mathbf{y} = \lambda_1^n\left|b_1\mathbf{x}_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^n b_2\mathbf{x}_2 + \left(\frac{\lambda_3}{\lambda_1}\right)^n b_3\mathbf{x}_3\right|$.

Furthermore, since $\lambda_1 > \lambda_2, \lambda_3, \ldots,$ and $b_1 \neq \mathbf{0}$, it followed that $\lim_{n\to\infty} \mathbf{A}^n\mathbf{y} = \lambda_1^n b_1 \mathbf{x}_1,$ so repeated application of the matrix to an arbitrary vector in the sampled *S. damnosum s.l.* riverine larval habitat dataset resulted in a vector proportional to the eigenvector with largest eigenvalue.

We then determined probabilities for the autoregressive model. Thereafter, we used $\frac{dS_t}{S_t} = \mu' dt + \sigma dW_t,$ where W_t was a P-standard Brownian motion (i.e., a standard Brownian motion under the probability measure P). We generated: $\frac{dS_t}{S_t} = (r - g)dt + \sigma\underbrace{(dW_t + \lambda dt)}_{=\hat{W}}.$ Then we used the Black and Scholes model whereby, the Q-martingale property was transferred to the value of predictor variable C in the *S. damnosum s.l.* larval habitat distribution model. Since the Girsanov theorem states that there is a probability measure Q such that $\hat{W}$ is a Q-standard Brownian motion and $e^{-rt}e^{gt}S_t$ and $e^{-rt}e^{gt}M_t$ are Q-martingales [8], in this research, $C(t)e^{-rt} = E_Q\left[C(T)e^{-rt}|F_t\right] \Rightarrow C(t) = e^{-r(T-t)}E_Q\left[(S(T) - k)^+|F_t\right].$ We then defined the S set by: $E = \{\omega \in \Omega | S(T)(\omega) \geq k\}F_t],$ which rendered $C(t) = e^{-r(T-t)}E_Q[S_T I_E|F_t] - ke^{-r(T-t)}E_Q[I_E|F_t].$ This quantity was computed by splitting each of its terms. The second term in the model generated $E_Q[I_E|F_t] = P_Q(E) = P_Q(S_T \geq k|F_t)$ using $S(T) = S(t)\exp\left\{(r - \frac{\sigma^2}{2})(T-t) + \sigma\hat{W}(T-t)\right\}.$ We then employed the condition $S_T \geq k$ to quantitate the riverine larval habitat parameter estimators which then defined $Y = -\frac{\hat{W}(T-t)}{\sqrt{T-t}} \leq d_2.$ The properties of the Brownian motions allowed us to write the expression $Y \sim N(0, 1)$ for the *S. damnosum s.l.* larval habitat distribution model, which then rendered $P_Q(S_T \geq k) = P_Q(Y \leq d_2) = N(d_2).$ Using the first

term $E_Q[S_T I_E] = \int_k^\infty x f_{S_T}(x)dx$, we were able to subsequently generate $S(T) = S(t)\exp\left\{(r - \frac{\sigma^2}{2})(T-t) + \sigma \hat{W}(T-t)\right\}$. The log-normal property of the underlying motion then rendered: $L_Q((r - \frac{\sigma^2}{2})(T-t) + \sigma \hat{W}(T-t))$ $= N((r - \frac{\sigma^2}{2})(T-t), \sigma\sqrt{T-t})$ and $e^{-r(T-t)}E_Q[S_T I_E|F_t] = S(t)N(d_1)$.

We then generated the inverse-Gamma distribution which in this research was a univariate specialization of the inverse-Wishart distribution summarized by using the regressed georeferenced *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-related seasonal-sampled explanatory covariates. The pdf was then $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta}{x}\right)$, while the mean of the model was $\frac{\beta}{\alpha - 1}$ for $\alpha > 1$. The variance was $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$. The skewness was $\frac{4\sqrt{\alpha-2}}{\alpha-3}$ for $\alpha > 3$, while the kurtosis was $\frac{30\alpha - 66}{(\alpha-3)(\alpha-4)}$ for $\alpha > 4$ and the entropy was $\alpha + \ln(\beta\Gamma(\alpha)) - (1+\alpha)\Psi(\alpha)$. The moment generating function was $\frac{2(-\beta t)^{\frac{\alpha}{2}}}{\Gamma(\alpha)} K_\alpha\left(\sqrt{-4\beta t}\right)$, while the characteristic function was $\frac{2(-i\beta t)^{\frac{\alpha}{2}}}{\Gamma(\alpha)} K_\alpha\left(\sqrt{-4i\beta t}\right)$.

The model revealed that when $p = 1$ (i.e., univariate) and $\alpha = m/2$, $\dot{\beta} = \Psi/2$, and $x = B$ and the pdf of the inverse-Wishart distribution became $p(x|\alpha, \beta) = \frac{\beta^\alpha x^{-\alpha-1}\exp(-\beta/x)}{\Gamma_1(\alpha)}$. The pdf of the Gamma distribution was $f(x) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k \Gamma(k)}$. We then defined the transformation $Y = g(X) = \frac{1}{X}$, employing the resulting transformation where we found: $f_Y(y) = f_X$

$\left(g^{-1}(y)\right)\left|\frac{d}{dy}\,g^{-1}(y)\right| = \frac{1}{\theta^k\Gamma(k)}\left(\frac{1}{y}\right)^{k-1}\exp\left(\frac{-1}{\theta y}\right)\frac{1}{y^2} = \frac{1}{\theta^k\Gamma(k)}\left(\frac{1}{y}\right)^{k+1}\exp\left(\frac{-1}{\theta y}\right) = \frac{1}{\theta^k\Gamma(k)}\,y^{-k-1}\exp\left(\frac{-1}{\theta y}\right)$. Replacing k with α; θ^{-1} with β; and y with x the resulted in the inverse-Gamma pdf shown above: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp\left(\frac{-\beta}{x}\right)$. The inverse Gamma distribution's pdf was then defined over the support $x > 0$ using the equation $f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(x)^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right)$, with shape parameter α and scale parameter β. The cumulative distribution function was then quantified by using the regularized Gamma function $F(x;\alpha,\beta) = \frac{\Gamma\left(\alpha,\frac{\beta}{x}\right)}{\Gamma(\alpha)} = Q\left(\alpha,\frac{\beta}{x}\right)$, where the numerator in the larval habitat model was the upper incomplete Gamma function and the denominator was the Gamma function. The regularized Gamma functions was then defined by $\frac{\gamma(a,\,z)}{\Gamma(a)}$ and $\frac{\Gamma(a,\,z)}{\Gamma(a)}$, where $\gamma(a,\,z)$ and $\Gamma(a,\,z)$ were incomplete Gamma functions and $\Gamma(a)$ was complete Gamma function.

We then employed methods in SASmacro WinBUGSio to calculate the multivariate Gamma function for the *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk model. This was constructed using $\Gamma_p(a) = \int_{S>0}\exp(-\text{trace}(S))|S|^{a-(p+1)/2}dS$, where $S > 0$ thus, S was positive-definite. We used the Gamma function to determine the recursive relationships in the sampled georeferenced larval habitat endemic transmission oriented explanatory covariates using $\Gamma_p(a) = \pi^{(p-1)/2}\Gamma(a)\Gamma_{p-1}(a-\frac{1}{2}) = \pi^{(p-1)/2}\Gamma_{p-1}(a)\Gamma[a+(1-p)/2]$. Thereafter, we quantified $\Gamma_1(a) = \Gamma(a)$, $\Gamma_2(a) = \pi^{1/2}\Gamma(a)\Gamma(a-1/2)$, and $\Gamma_3(a) = \pi^{3/2}\Gamma(a)\Gamma(a-1/2)\Gamma(a-1)$. We then defined the multivariate digamma function in the larval habitat model as $\psi_p(a) = \frac{\partial\log\Gamma_p(a)}{\partial a} = \sum_{i=1}^{p}\psi(a+(1-i)/2)$ and the general polygamma

function as $\psi_p^{(n)}(a) = \dfrac{\partial^n \log \Gamma_p(a)}{\partial a^n} = \sum_{i=1}^{p} \psi^{(n)}(a + (1 - i)/2)$. The digamma function in the *S. damnosum s.l.* larval habitat model was then defined as the logarithmic derivative of the Gamma function: $\psi(x) = \dfrac{d}{dx} \ln \Gamma(x) = \dfrac{\Gamma'(x)}{\Gamma(x)}$. This equation then calculated the digamma function which in this research was expressed as $\dfrac{\partial \Gamma(a + (1 - i)/2)}{\partial a} = \psi(a + (i - 1)/2)\Gamma(a + (i - 1)/2)$. We then generated the following expression:

$$\frac{\partial \Gamma_p(a)}{\partial a} = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(a + (1 - j)/2) \sum_{i=1}^{p} \psi(a + (1 - i)/2) = \Gamma_p(a) \sum_{i=1}^{p} \psi(a + (1 - i)/2).$$

Since in this research $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(a + \dfrac{1 - j}{2})$, it followed that $\dfrac{\partial \Gamma_p(a)}{\partial a} = \pi^{p(p-1)/4} \sum_{i=1}^{p} \dfrac{\partial \Gamma(a + \frac{1-i}{2})}{\partial a} \prod_{j=1, j \neq i}^{p} \Gamma(a + \dfrac{1 - j}{2})$. The pdf of the inverse Wishart for the riverine larval habitat model was found to be $\dfrac{|\Psi|^{\frac{\nu}{2}}}{2^{\frac{\nu p}{2}} \Gamma_p\left(\frac{\nu}{2}\right)} |\boldsymbol{X}|^{-\frac{\nu+p+1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\Psi \boldsymbol{X}^{-1})}$, when $\mathbf{X}$ and Ψ were $p \times p$ positive definite matrices, and $\Gamma_p(\cdot)$ was the multivariate Gamma function.

In this research, when *f* was convex in the *S. damnosum s.l.* riverine larval habitat model and twice continuously differentiable and the sublevel set $\{x : f(x) \leq f(x0)\}$ was bounded (e.g., seasonal-sampled georeferenced larval habitat), then the sequence of function values generated by the BFGS method with an inexact Armijo-Wolfe line search converged at the minimal value (i.e., larval density count) of *f*. This result did not follow directly from the standard Zoutendijk theorem as commonly quasi-Newton methods do, but instead the eigenvalues of the inverse Hessian approximation *Hk* did not grow too large or too small. If the convexity assumption is dropped, pathological counter examples to convergence are known to exist [2]. According to Booth and Hobert [10], if $xk + 1 = xk + \_\, kpk$ is an iteration, where $\_k$ satisfies the Wolfe

condition, then the Zoutendijk theorem states that $f(x)$ will be bounded below, while still be continuously differentiable, which then can be expressed as $\{x \mid f(x) \le f(x0)\}$, where ∇f is Lipschitz continuous, i.e., $\forall x,\ y \in N \quad k\nabla f(x) - \nabla f(y)k \le Lkx - yk.$ In our model this procedure rendered $Xk\_0\cos 2\_kk\nabla \quad fkk2 < \infty.$

The search direction $\boldsymbol{p}_k$ at stage k was then given by the solution of the analogue of the Newton equation $\boldsymbol{B}_k\boldsymbol{p}_k = -\nabla f(\mathbf{x}_k)$, where $\boldsymbol{B}_{kn}$ was an approximation to the Hessian matrix, which was updated iteratively at each stage in the seasonal predictive *S. damnosum s.l.* riverine larval habitat model when $\nabla f(\mathbf{x}_k)$ was the gradient of the function evaluated at $\mathbf{x}_k$. A line search in the direction $\boldsymbol{p}_k$ was then employed to find the next point $\mathbf{x}_{k+1}$. Instead of requiring the full Hessian matrix at the point $\mathbf{x}_{k+1}$ to be computed as $\boldsymbol{B}_{k+1}$, the approximate Hessian at stage k was updated by the addition of two matrices. $\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \boldsymbol{U}_k + \boldsymbol{V}_k$. Both U_k and V_k were then symmetric rank-one matrices but with different matrix bases. The symmetric rank one assumption thus meant that we could write $C = ab^T$. U_k and V_k which in this research we employed as a rank-two update matrix. We found that our model was robust against the scale problem often suffered in the gradient descent searching (e.g., in Broyden's method). The quasi-Newton condition imposed on this update was $\boldsymbol{B}_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$.

From an initial guess $\boldsymbol{X}_0$ and an approximate Hessian matrix $\boldsymbol{B}_0$, estimators were repeated until $\boldsymbol{x}$ converged to solution. We obtained a direction $\boldsymbol{p}_k$ by solving: $\boldsymbol{B}_k\boldsymbol{p}_k = -\nabla f(\mathbf{x}_k)$. We then performed a line search to find an acceptable step size α_k in the direction found in the first step, which then updated $\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \alpha_k\boldsymbol{p}_k$. We then set $\boldsymbol{s}_k = \alpha_k\boldsymbol{p}_k$, and solved for $\boldsymbol{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$, $B_{k+1} = B_k + \dfrac{y_k y_k^T}{y_k^T s_k} - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}$. The $f(\mathbf{x})$ denoted the objective function to be minimized. Convergence can be checked by observing the norm of the gradient, $|\nabla f(\mathbf{x}_k)|$.

Practically, $\boldsymbol{B}_0$ can be initialized with $\boldsymbol{B}_0 = \boldsymbol{I} * x$, so that the first step will be equivalent to a gradient descent, but further steps are more and more refined by $\boldsymbol{B}_k$, using the approximation to the Hessian [1]. The first step of the algorithm was thereafter carried out using the inverse of the matrix $\boldsymbol{B}_k$, which in this research was obtained efficiently by applying the Sherman- Morrison formula to the fifth line of the algorithm, thus rendering $\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{B}_k^{-1} + \frac{(\mathbf{s}_k^T \boldsymbol{y}_k + \boldsymbol{y}_k^T \boldsymbol{B}_k^{-1} \boldsymbol{y}_k)(\mathbf{s}_k \mathbf{s}_k^T)}{(\mathbf{s}_k^T \boldsymbol{y}_k)^2} - \frac{\boldsymbol{B}_k^{-1} \boldsymbol{y}_k \mathbf{s}_k^T + \mathbf{s}_k \boldsymbol{y}_k^T \boldsymbol{B}_k^{-1}}{\mathbf{s}_k^T \boldsymbol{y}_k}$. The Sherman-Morrison formula is a formula that allows a perturbed matrix to be computed for a change to a given matrix **A**. If the change can be written in the form $\mathbf{u} \otimes \mathbf{v}$ for two vectors **u** and **v**, then the Sherman-Morrison formula for the two vectors **u** and **v**, using the Sherman-Morrison formula is $(\mathbf{A} + \mathbf{u} \otimes \mathbf{v})^{-1} = \mathbf{A}^{-1} - \frac{\left(\mathbf{A}^{-1}\mathbf{u}\right) \otimes \left(\mathbf{v} \cdot \mathbf{A}^{-1}\right)}{1 + \lambda}$ [7]. In this research, for the two vectors **u** and **v**, the Sherman-Morrison formula was $(\mathbf{A} + \mathbf{u} \otimes \mathbf{v})^{-1} = \mathbf{A}^{-1} - \frac{\left(\mathbf{A}^{-1}\mathbf{u}\right) \otimes \left(\mathbf{v} \cdot \mathbf{A}^{-1}\right)}{1 + \lambda}$, where $\lambda \equiv \mathbf{v} \cdot \mathbf{A}^{-1}\mathbf{u}$. In statistical estimation problems (such as ML or Bayesian inference), credible intervals or confidence intervals for the solution can be estimated from the inverse of the final Hessian matrix [2]. However, in our *S. damnosum s.l.* riverine larval habitat model these intervals were technically defined by the true Hessian matrix, as such, BFGS approximation did converge to the true Hessian matrix.

Table 6 presents the results of the Poisson regression for the interactions model. These results provided information for estimates of the prior distribution based on the regressed main effect coefficients for the Bayesian analysis. The values for parameter estimates and standard errors in Table 6 were then used as mean values and standard errors to parameterize prior expected values for the sampled riverine larval habitat endemic transmission oriented explanatory covariates. The prior expected mean value for the error term was assumed to be zero (0), with a standard deviation of 0.01. Initial values for the MCMC chains were then generated. Three MCMC chains were then estimated for the

intercept, which appeared to converge within the first 1,000 samples. The first 1,000 samples were discarded to allow the model to stabilize (i.e., known as "burn in"), and the next 10,000 samples were used to derive parameter estimates. The MCMC was able to numerically calculate multi-dimensional integrals. In our MCMC methods, an ensemble of "walkers" moved around randomly. At each point where the walker stepped, the integrand value at that point was counted towards the integral. The walker then made a number of tentative steps around the area looking for a place with reasonably high contribution to the integral. Random walk methods are a kind of random simulation or Monte Carlo method [1]. However, whereas standardized random samples of the integrand use a conventional Monte Carlo integration, which are generally statistically independent those used in our MCMC were correlated. A Markov chain was then constructed in such a way as to have the integrand as its equilibrium distribution. Table 6 lists the improvement in model fit, as variables were added to the Bayesian model.

Table 6. Results of SAS Poisson regression used to estimate prior distribution of coefficients for the MCMC analysis

Variable	df	Coefficient	SE	*P*
Intercept	1	1.4134	0.1264	< 0.0001
AQVEG	1	0.03943	0.0061	0.4332
HGVEG	1	0.02653	0.0049	< 0.0001
DDVEG	1	0.0118	0.0131	< 0.0001
MMB	1	0.0542	0.1698	0.7176

This specification moved the investigation towards a Bayesian map analysis, given that the entire *S. damnosum s.l.* larval habitat endemic transmission-oriented predictive risk-based explanatory covariates, with the exception of the intercept, were treated as single-valued. In this research, the intercept was treated as a distribution of values and was estimated by using empirical Bayes techniques.

Next, the difference in the deviances between a simple model and the more complex model provided the improvement χ^2 values listed in Table 6. We examined all interaction between the sampled georeferenced *S. damnosum s.l.* larval habitat covariates and found that an interaction model did not improve the fit therefore, no interaction terms were included in the final model. We could not examine the improvement of fit between a saturated model and the full effects model as the number of the sampled parameters that needed to be estimated exceeded the maximum number that could be parsimoniously regressed. To derive the improvement of fit values listed in Table 7, the posterior mean deviance values were obtained with deviance information criterion (DIC) spatial analytical tools. We focused on a spatial consideration of the local DIC measure for model selection and goodness-of-fit evaluation. We employed a partitioning of the DIC into the local DIC, leverage, and deviance residuals to assess the local model fit and influence of the sampled *S. damnosum s.l.* riverine larval habitat observations in the Bayesian framework. We also used visualization of the local DIC to assist in model selection and to visualize the global and local impacts of adding covariates or model parameters to the Bayesian estimation matrix. DIC statistics were then generated to identify the best fitting model.

In this research, the deviance was defined as -2^* log (likelihood), where the 'likelihood' was defined as $p(y \mid$ and theta). This included all the normalizing constants where y comprised all stochastic node values and theta stochastic parents of y. 'Stochastic parents' are the stochastic nodes upon which the distribution of y depends upon when collapsing over all logical relationships [8]. For example, in this research, $y \sim$ Dnorm (mu, tau), then tau was a function of a parameter phi over which the prior distribution has been placed. As such, then the likelihood was defined as a function of phi of our *S. damnosum s.l.* riverine larval habitat model. The expectation $\overline{D} = E^{\theta}[D(\theta)]$ was then used as a measure of model fitness based on the values of the sampled endemic transmission oriented covariate coefficient values. The effective number

of parameters included in the model was computed as $p_D = \overline{D} - D(\overline{\theta})$, where $\overline{\theta}$ was the expectation of θ. The DIC then generated the following conclusions: (1) the Dbar, was the posterior mean of the deviance, (2) the Dhat, was the point estimate of the deviance (i.e., -2^* log (likelihood)) obtained by substituting the posterior means theta bar of theta, which then rendered Dhat $= -2^*$ log ($p(y$ — theta. bar); and, 3) pD was the effective number of riverine larval habitat parameters provided by pD = Dbar – Dha and pD employing the posterior mean of the deviance minus the deviance of the posterior means. In normal hierarchical models, Pad = TR (H), where H is the 'hat' matrix maps the observed data to their fitted values [8]. The DIC was then calculated as: $DIC = p_D + D$. The DIC value for the final model was 931.6.

Table 7. Improvement of fit of the WinBUGS hierarchical Bayesian model (HBM) *S. damnosum s.l.* larval habitat model

Unadjusted effects			Adjusted effects	
	Improvement		Improvement	
Variable	df	χ^2	χ^2	df
HGVEG	1	1.133	1.494	1

Median parameter values, as well as the 95% credibility intervals (2.5 percentile and 97.5 percentile values), are listed in Table 8. As the *S. damnosum s.l.* riverine larval habitat sampling sites increased based on the sampled georeferenced explanatory covariate distance percent of hanging vegetation, the median log-count of larval count increased. The adjusted model that assumed independence among the field and remote-sampled endemic transmission oriented explanatory covariates of the larval counts fit better that the model that adjusted for correlation within the Nabere study site based on the RMSE.

4. Discussion

In this research, an SAS-based cluster analyses helped spatially identify high and low ABR-stratified clusters. The CLUSTER procedure hierarchically aggregated the sampled *S. damnosum s.l.* larval habitat observations in an SAS dataset. Although PROC CLUSTER aggregated the explanatory predictor variables by georeferenced epidemiological sampled location site data, the SAS analyses did not identify the varying or constant cluster-based endemic transmission oriented covariates in the sampled data related to the ABR stratified clusters. Unfortunately, ordinary significance tests such as, analysis of variance *F*-tests are not possible to be derived from such data. There are no completely satisfactory methods for determining cluster-based varying and constant explanatory covariates within a population cluster in any type of analysis [1].

Importantly, since our SAS clustering method attempted to maximize the separation between clusters, the assumptions of the usual significance tests, parametric or nonparametric in the seasonal *S. damnosum s.l.* larval habitat-distribution model would have been drastically violated further nullifying identification and quantitation of the seasonal sampled cluster-based georeferenced explanatory covariates. For the same reason methods that purport to test for clusters against the null hypothesis where objects are assigned randomly to clusters such as in Booth and Hobert [13] would be useless for identifying varying and constant cluster-based endemic transmission oriented explanatory predictor covariates of productive *S. damnosum s.l.* riverine larval habitats. Most valid tests for spatial clusters either have intractable sampling distributions or involve null hypotheses for which rejection is uninformative [1]. As such, hierarchical SAS-based cluster analyses cannot identify any seasonal-sampled georeferenced varying or constant *S. damnosum s.l.* larval habitat endemic transmission-oriented explanatory covariate coefficients associated to productive habitats based on field-sampled count data.

We then constructed a Poisson model in PROC NL MIXED as we assumed the response variable in the analyses of the sampled *S. damnosum s.l.* riverine larval habitat data had a Poisson distribution. We assumed the logarithm of the expected value could be modeled by a linear combination of the ecological-sampled parameters. In this research, the variance function was dependent on the iteration's estimate of the model parameters; however, there was overdispersion present in the model. Thus, we expressed the variance of the response as negative binomial model, whereby $(y) = \mu + k\mu^2$. The variance of an overdispersed Poisson model is a linear function of the mean, while the variance of a negative binomial model is a quadratic function of the mean [1]. Because the GLM employed in risk-based riverine larval habitat data analyses is based on ML or quasi-likelihood, the model may be very sensitive to spurious observations thus, generating overdispersion in residually forecasted estimators [11].

In this research, the negative binomial riverine larval habitat models with mean $\mu_i + \alpha\mu_i^P$, generated where in general $-\infty < p < \infty$. NL MIXED estimated two negative binomial models, corresponding to $p = 2$ with variance $\mu_i + \alpha\mu_i^2$ and $p = 1$ with variance $\mu_i + \alpha\mu_i$. For Poisson data, the variance depends on the mean, and, thus on the model parameters [1]; hence, the geographically weighted least-squares equations, we used for our regression had to be weighted by a diagonal variance matrix based on the variance function calculated for the sampled larval habitat data. The first model was estimated with the option DIST=NEGBIN $(p = 2)$, and the second model was estimated by using DIST=NEGBIN $(p = 1)$. Because the variance is a function of the mean, large and small *S. damnosum s.l.* riverine larval habitat counts got weighted differently in the negative binomial regression matrix. Fortunately, this approach was efficient for analyzing the sampled riverine larval habitat parameters as the target value μ was modeled directly making inference straightforward while avoiding the need of data re-transformation. Variance relationships affect the weights in the iteratively reweighted least squares algorithm for fitting models to data

[1]. Moreover, the model enabled us to go beyond the location-scale family considered in the previously published literature and allowed some flexibility through the choice of the link function (i.e., logarithmic, inverse), and, of the riverine larval habitat regression-based distribution of *Y* through its expectation-variance relationship.

In this research, NL MIXED procedure provided a flexible environment in which to model the seasonal-sampled *S. damnosum s.l.* riverine larval habitat data in time for accurately quantitating correlations among the error measurements. This was performed by employing random effects estimates and random regression coefficients. Commonly, GLM uses the OLS estimation in spatiotemporal predictive seasonal vector arthropod-related larval habitat distribution models that to quantitate parameter estimate values, to minimize the squared difference between observed and predicted values of the dependent variable (e.g., larval habitat count data) which then are subsequently utilized to attain pseudo R^2 values [8]. In this research, this approach led to the familiar analysis of variance table in which, the heterogeneity of the sampled *S. damnosum s.l.* endemic transmission-oriented predictive risk-based data quantitated the total sum of squares by dividing the specific variabilities in the regressed data. However, PROC NL MIXED did not produce an analysis of variance table. Instead, the procedure generated REML estimators. REML method is a variant of ML estimation; whereby, estimators are obtained not from maximizing the whole likelihood function but, only that part that is invariant to the fixed effects part of the linear model [1]. Thereafter, the spatiotemporal *S. damnosum s.l.* riverine larval habitat distribution model was generated employing, $y = Xb + Zu + e$, where Xb was the fixed effects in the sampled explanatory predictor covariate coefficients, Zu was the random effects part, and e was the error term. The REML estimates were then obtained by maximizing the likelihood function of $K'y$, where K was a full rank matrix with columns orthogonal to the columns of the X matrix (i.e., $K'X = 0$). Thereafter, the uncertainty correlation estimation determined the REML as the estimator of the variance-covariance matrix of y (e.g., V). Fortunately, this did not depend on the choice of matrix K since the GLS equations rendered from the weighted

least squares approach and the GLM procedure in the larval habitat distribution model became X'(inverse of V) $Xb = X'$ (inverse of V) y, where V was replaced with its estimator which was then subsequently solved to obtain the estimates of fixed effects parameters b in the model. We noted that when u and e in our riverine larval habitat distribution model were not correlated and when V (i.e., the variance-covariance matrix of y) was equal to $ZGZ'+R$, then G and R were the variance matrices of u and e, respectively.

Additionally, a PRIOR statement to PROC NL MIXED code enabled constructing a sampling based Bayesian probabilistic estimation matrix from a variance-component model. In the Bayesian matrix, we were able to instantiate the nodes directly which was equivalent to supplying a measurement for a radius based on the distribution of the georeferenced riverine larval habitat estimators. We were able to quantitate an uncertainty estimate in our likelihood information by supplying a probability distribution over a set of radii generated from the sampled data. We used a backward propagation to compute the node probabilities. The error gradients were then propagated in any direction throughout the Bayesian hierarchical regression-based framework. The probability propagation allowed choosing between the sampled *S. damnosum s.l.* georeferenced endemic transmission oriented explanatory covariates using the calculated probability of each sampled covariate coefficient value based on one or more of the nodes. Our Bayesian approach allowed flexible model fitting and estimation of all "high risk" riverine larval habitats based on the seasonal-sampled larval count data. The algorithm produced a sequence of parameter vectors that represented random draws from the posterior distribution. Our results indicated that likelihood weights influenced the resulting posterior distributions of the seasonally quantitated field and remote-sampled *S. damnosum s.l.* larval habitat parameter estimators, which, in turn, influenced summarizing the spatial trends in the variance uncertainty estimates for accurate model prediction. By doing so, the sampled parameter estimator Percentage of hanging vegetation was found to be the most important covariate associated to the spatiotemporal-sampled riverine larval habitats at the Nabere epidemiological riverine study site.

A spatial residual trend analyses was then performed by using the uncertainty indices, which linked tabular data with the sampled count data in ArcGIS®. Thereafter, the estimation matrix identified prolific riverine habitats based on the georeferenced regressed explanatory covariate coefficients. Similium larval stages are commonly found in running water where rocks break the water surface and the turbulence of the water results in a higher level of oxygenation [2]. The main factors governing *S. damnosum s.l.* larval habitats are adequate water velocity (0.70-1.5m/sec), which is linked with oxygenation and food supply and the presence of suitable supports which may be vegetation rocks, stones, sills, sidewalks of structures, spillways, and gates [12]. All the seasonal sampled empirical data was then exported into SAS/GIS®. The SAS/GIS® Batch Import process allowed us to employ SAS® Component Language (SCL) code to import the regression models and the ArcGIS-derived predictors. The SAS/GIS® Batch Import process allowed us to define the sampled *S. damnosum s.l.* riverine larval habitat parameter values that were needed for the import through macro variables and SAS® file. Thereafter, we defined and quantitated the seasonal-sampled larval density count values. We called an SCL entry to actually initiate the import. The process included specification of the sampled riverine larval habitat endemic transmission-oriented predictive risk-based explanatory covariate coefficients to import.

In this research, we defined the *S. damnosum s.l.* larval habitat input parameters by setting the values of the macro variable and by assigning filerefs. We then executed the SASHELP.GISIMP.BATCH.SCL entry to start the Batch Import process. We did not pass any of the sampled georeferenced endemic transmission oriented explanatory covariates directly to the SCL entry; larval habitat parameters had to be defined through macro variables and filerefs before we called the SCL entry. In this research, we specified the sampled riverine larval habitat parameter estimators measurement indicator values for the import by assigning a fileref, using the SYMPUT/SYMPUTN function in SCL. The importing window interface let us modify the default composites and the default layer definitions before we proceeded with the import. The SASHELP.GISIMP dataset supplied the covariate coefficient values that

we needed to import as well. Included in this dataset were two variables, DEFMLIB and DEFSLIB, which were used to supply the default values for the map entries library and the spatial data library. To modify the composites and layers before the import occurred, we used the importing window; however, after employing the importing window, we also used PROC GIS to make changes to the seasonal *S. damnosum s.l.* larval habitat distribution model and its underlying components.

To open the GIS Spatial Data Importing window, we had to assign the sampled habitat location of the GISIMP dataset to the macro variable USER_FIL. In SAS/GIS software, the *S. damnosum s.l.* riverine larval habitat attribute data was stored in SAS view. As such we created an SAS/ACCESS view to access data in a database (i.e., DB2). By doing so, a DATA step view then enabled us to access an external file for any seasonal sampled georeferenced riverine larval habitat endemic transmission-oriented predictor variable. Once our attribute data was accessible through the SAS view, it was linked to the explanatory covariates, which then helped in labelling (i.e., theming) the sampled georeferenced data. For example, our remote data represented a portion of the riverine study site and contained information, such as sampled georeferenced larval habitat boundaries, surrounding vegetation land cover and so on. An attribute data set with population information for each census tract was then linked in SAS/GIS by using the corresponding tract composite in the spatial data.

A semiparametric filtering model then employed proxy variables constructed from the sampled *S. damnosum s.l.* larval habitat explanatory covariates to replace the misspecification terms for implementing a conditional covariance matrix employing the autoregressive specification in SAS/GIS. An unknown misspecification term can be approximated by a set of spatial proxy variables [1]. After quantitating the misspecification terms in the riverine larval habitat distribution model, the remaining residuals $\hat{\varepsilon}$ became white noise. This result implied that the estimated regression parameters $\hat{\beta}$ were unbiased for the basic regression model, $y = X\beta + \varepsilon^*$, where ε^* incorporated the misspecification term and the white-noise disturbances.

This allowed us to calibrate the autoregressive models with the standard OLS estimation procedure. By doing so, an SAS dataset containing a pseudo-random sample from the joint posterior density of the variance-components model and fixed effects in a mixed model was then able to exploit the spatiotemporal remotely-sampled immature *S. damnosum s.l.* data.

Thereafter, a posterior sample was generated, which resulted from an MIVQUEO fit, which was then used to create the sample in the model. By default SAS used an independence chain algorithm in order to generate the sample. This algorithm worked by creating a pseudo-random proposal from a convenient bare distribution of the spatiotemporal-sampled georeferenced explanatory covariate coefficient estimates, which was then chosen to be as close as possible to the posterior in the riverine larval habitat model. Then the proposal retained the sampled *S. damnosum s.l.* endemic transmission-oriented predictive risk-based endemic transmission oriented explanatory covariate coefficient indicator values with probability proportional to the ratio of the weights constructed by taking the ratio of the true posterior to the base density. If a proposal is not accepted, then a duplicate of the previous observation can be added to the chain [7]. This algorithm was then customized for the mixed model. The state of the chain was then used as a sample of the desired distribution for the spatiotemporal data analyses. These sequences were then used to approximate the riverine larval habitat distribution (i.e., to generate a histogram), and to compute integrals (i.e., expected predictor values).

One of the most important parts of our eigendecomposition of the *S. damnosum s.l.* riverine larval habitats explanatory covariates was the Jacobian matrix. For example, in the Jacobian matrix, if the function $F\,I$ was differentiable at a sampled larval habitat point $p = (x_1, \ldots, x_n)$, then the derivative of F at p was the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ represented by the matrix $J_F(x_1, \ldots, x_n)$. In this research, this linear transformation was the best linear approximation of the function F near the sampled larval habitat point p at the study site. Further, the Jacobian matrix for our larval habitat model was a square matrix, and its

determinant was a function of $x_1, \ldots, x_n$, which in this research was the Jacobian determinant of *F*. As such, the predictive seasonal arthropod-related risk model contained important information about the local behaviour of *F* and thus, was a local expansion factor for volumes in the residual forecasts targeting the endemic transmission oriented larval habitat associated explanatory covariates.

Therefore, a robust Jacobian matrix may be employed when performing variable substitutions in multivariable integrals for representing seasonal sampled vector arthropod-related explanatory covariate coefficients since it would occur prominently in the substitution rule for seasonal sampled larval habitat variables. For example, if a vector ecologist or a local abatement district manger employs the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ in a predictive *S. damnosum s.l.* larval habitat model employing $F(x, y) = \begin{bmatrix} x^2 y \\ 5x + \sin(y) \end{bmatrix}$, then the residual forecasts targeting the endemic transmission zones would reveal $F_1(x, y) = x^2 y$ and $F_2(x, y) = 5x + \sin(y)$ and the Jacobian matrix of *F* would be $J_F(x, y) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 5 & \cos(y) \end{bmatrix}$. Thereafter, the Jacobian determinant would be represented by $\det(J_F(x, y)) = 2xy\cos(y) - 5x^2$. By doing so, the Jacobian would then generalize the gradient of a scalar-valued function of the seasonal-sampled immature *S. damnosum s.l.* predictive sampled variables which in turn, would generalize the derivative of a scalar-valued function of a single sampled variable. In other words, the Jacobian for a scalar-valued multivariable function in the predictive seasonal riverine larval habitat endemic transmission-oriented seasonal risk model would be the gradient and a scalar-valued function of single variable would be its derivative. Further, the Jacobian in the predictive model would be able to describe the amount of "stretching", "rotating" or "transforming" that log- a transformation imposes locally. For example, if $(x_2, y_2) = f(x_1, y_1)$ is employed to

transform a georeferenced *S. damnosum s.l.* riverine larval habitat image, the Jacobian of $J(x_1, y_1)$ would then describe how the image in the neighbourhood of (x_1, y_1) is transformed.

Therefore, an autoregressive process in the error term may then also be employed to a seasonal *S. damnosum s.l.* riverine larval habitat model to derive the sample distribution of the Moran's I statistic for quantitating latent autocorrelation components in a dataset of residual forecasts derived from an eigenfunction decomposition algorithm. The spatial filter eigenvectors can then establish means, variances, distributional functions, and pairwise correlations for statistically targeting significant seasonal-sampled endemic transmission oriented predictor variables. The redundant information resulting from any spatial spill-overs, will then be seasonally quantitated in the mean response term of the predictive risk model as a linear combination of various distinct seasonal riverine larval habitat map patterns. Additionally, the fixed-effects sampled parameter estimates in a remotely-sensed *S. damnosum s.l.* riverine larval habitat model can be analytically integrated out of the joint posterior leaving the marginal posterior density of the variance components. In order to better approximate the marginal posterior density of the variance, the spatiotemporal-sampled immature riverine larval habitat endemic transmission-oriented predictive risk data can then be further transformed by using MIVQUE(0) equations. Thereafter, the density of the transformed data can be approximated by using Bayesian statistics.

In this research, Bayesian regression equations were employed to quantitate the inter-relationship between QuickBird derived seasonal *S. damnosum s.l.* riverine habitats endemic transmission oriented covariate coefficients. A PRIOR statement to PROC NL MIXED code was initially employed to construct a Bayesian probabilistic estimation matrix from a variance-component model. Bayesian probability is one of the different interpretations of the concept of probability and belongs to the category of evidential probabilities which is essentially an extension of logic that enables reasoning with uncertain statements for evaluating a hypothesis related to residual uncertainty quantitation [2]. The design of

the *S. damnosum s.l.* mixed model in this research, was constructed in such a manner as to specifically incorporate residual autocorrelation error components while including the influence of other georeferenced a spatial predictor variables in the QuickBird sampled data. Spatially lagged and simultaneous autoregressive models were also built based on multiple remote-sampled endemic transmission oriented explanatory covariate coefficients. The coefficient estimates were then used to define expectations for prior distributions in the Bayesian matrix by using MCMC specifications. A seasonal spatial residual trend analyses was then performed using uncertainty indices, which then linked the tabular data with the sampled *S. damnosum s.l.* data in SAS/GIS. Thereafter, the estimation matrix identified prolific seasonal-sampled riverine larval habitats based on the georeferenced explanatory predictors in WinBUGS.

In this research, WinBUGSio macro provided functionality for the input and output of information to and from WinBUGS and SAS/GIS. The macro provided an optimal number of different options for specifying the seasonal-sampled *S. damnosum s.l.* riverine larval habitat predictor variables in the empirical-sampled dataset. For example, the sampled riverine larval habitat dataset was written out to an ASCII text file in column format and an S language list (…) format, so we could specify an external ASCII file containing the sampled larval habitat data. We were required to specify which sampled georeferenced endemic transmission oriented predictor variables in the WinBUGS data file we wished to regress. Further, we also had to specify labels to be employed in the Bayesian probabilistic matrix. Since WinBUGS is case-sensitive, we had to be careful to use labels which corresponded to the sampled data items specified in the *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented distribution model (e.g., covariate coefficient labels, response variable names etc). Fortunately, this step allowed using a new SAS dataset with the existing WinBUGS model without having to rename the sampled georeferenced riverine larval habitat predictor variables in the model. Thereafter, we wrote the batch code for running the model remotely in WinBUGS. The standard WinBUGS batch commands for checking model syntax then read the appropriate empirical data while simultaneously reading in value files pre-specified in macro.

We then specified the appropriate directory path and file names for the various files and a listing of endemic transmission-oriented the riverine larval habitat model parameters estimators to be sampled from the joint posterior distribution. This list created the necessary commands within the batch script to the sampled riverine larval habitat estimators. We noticed that WinBUGS required nodes to be monitored for the sampled data, so that the outputs could be read as summary statistics from the posterior distribution. Thereafter, the batch code commands for summary statistics from the posterior distribution were generated and the summary statistics were written out to a log file. The batch script was then written out to the default WinBUGS installation directory where we specified the file name for the script. The macro then executed this batch script in WinBUGS by executing a \\ command for running WinBUGS in batch mode. WinBUGS then opened the batch commands executed. Finally, the log file containing the summary statistics from the MCMC sample of the posterior distribution for each monitored node was read into the SAS dataset. The macro also provided functionality for us to select the sampled *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented risk-based estimators for all MCMC iterations stored (i.e., CODA output), which was then read in to SAS for post-processing. The WinBUGSio macro also specified a list of the riverine larval habitat model parameters in the summary statistics of the MCMC sample from the joint posterior distribution. These summary statistics included the mean, standard deviation, median, lower 5% and upper 95% quantiles, the number of samples and a measure of the Monte Carlo standard error of the mean, which in this research was provided by the sample standard deviation divided by the square root of the number of simulation draws (i.e., number of immature *S. damnosum s.l.* samples).

This research has laid down the foundation for robustly constructing more Bayesian paradigms for exploiting seasonal-sampled *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented explanatory covariate coefficients in future research. For example, SAS/STAT software now provides Bayesian analysis in downloadable, experimental versions of three procedures for SAS 9.2 on Windows: GENMOD, LIFEREG, and PHREG. The new BAYES statement in these procedures

can render various Bayesian probabilistic regressors employing georeferenced seasonal-sampled *S. damnosum s.l.* riverine habitat field-sampled data for determining seasonal predictive inference capabilities employing residuals from GLMs, Cox regression models, and piecewise constant baseline hazard models (i.e., piecewise exponential models). For example, Cox regression or proportional hazards regression would allow analyzing the effect of several seasonal *S. damnosum s.l.* endemic transmission-oriented explanatory covariate coefficients. The probability of the endpoint (e.g., recurrence of onchocerciasis in an epidemiological study site), for example, could then be classified as the hazard variable in a predictive endemic transmission-oriented risk model framework. The hazard in the regressed *S. damnosum s.l.* data could then be expressed as: $H(t) = H_0(t) \times \exp(b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k)$, where $X_1 \ldots \mathrm{X}_k$ would be the collection of spatiotemporal predictor variables and $H_0(t)$ would be the baseline hazard at time t, representing aggregations of sampled georeferenced larval habitats with varying density count covariate coefficient values. By dividing both sides of the equation by endemic transmission-oriented $H_0(t)$ and taking logarithms, the equation $\ln\left(\frac{H(t)}{H_0(t)}\right) = b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_kx_k$ can then be efficiently derived. In this predictive riverine larval habitat model $H(t)/H_0(t)$ would be the hazard ratio. The coefficients $b_i \ldots b_k$ would then be estimated by the Cox regression, and then interpreted in a similar manner to that of multiple logistic regression. For example, suppose a empirical datset of *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented explanatory covariate coefficients is dichotomous and is coded 1 if, present and 0 if, absent (e.g., presence of larvae). Then the quantity $\exp(b_i)$ can be interpreted as an instantaneous relative risk of a disease transmission event, at any time, for any seasonal-sampled covariate sampled at an epidemiological riverine study site. Thereafter, if the sampled *S. damnosum s.l.* riverine larval habiat explanatory covariate coefficient is continuous, then the quantity $\exp(b_i)$ would be the instantaneous relative risk of a disease transmission event, at any seasonal time frame as the residual forecasts targeting the endemic transmission zones would increase in the value

compared with another individual sampled covariates. These computations can then employ the Gibbs sampler to obtain a robust posterior distribution.

In statistics and in statistical physics, Gibbs sampling or a Gibbs sampler is an MCMC algorithm for obtaining a sequence of observations, which are approximately quantized from a specified multivariate probability distribution (i.e., from the joint probability distribution of two or more *S. damnosum s.l.* random variables), when direct sampling is difficult (e.g., flooded riverine ecosystems) [2]. It is a randomized algorithm (i.e., an algorithm that makes use of random numbers), and hence may produce different results each time it is run which can be an alternative to deterministic algorithms for generating robust statistical information from variational Bayes or EM algorithms. As with other MCMC algorithms, Gibbs sampling can generate a Markov chain based on seasonal-sampled georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented predictive risk-related explanatory covariate coefficients, each of which would be correlated with nearby samples. As a result, care must be taken if independent samples are desired by thinning the resulting chain of samples by only taking every n-th value (e.g., every 100th value). As in other MCMC algorithms, samples from the beginning of the chain (i.e., the burn-in period) may not accurately represent the desired distribution initially. Convergence diagnostics such as the Gelman-Rubin, Geweke, Heidelberger-Welch, and Raftery-Lewis tests however, can be produced as well as trace plots. All procedures offer the normal and uniform prior, and the BGENMOD procedure which can then provide a robust Jeffreys' prior for seasonally quantitating *S. damnosum s.l.* riverine larval habitat parameter estimators efficiently. In Bayesian probability, the Jeffreys prior is a non-informative (i.e., objective) prior distribution on parameter space that is proportional to the square root of the determinant of the Fisher information: $p(\vec{\theta}) \propto \sqrt{\det \mathcal{I}(\vec{\theta})}$. It has the key feature that it is invariant under reparameterization of the parameter vector $\vec{\theta}$. This makes it of special interest for use with scale parameters in seasonal multivariate predictive *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented model.

A vector ecologist or a local district-level abatement manager may also derive a robust output employing the posterior distribution to an SAS dataset for performing additional analysis (stochastic/deterministic interpolation) in SAS/GIS. Currently, the experimental BGENMOD, BLIFEREG, and BPHREG procedures are being made available through the SAS website so that users can generate robust residual forecast for various input for cartographic interpolation. These models can therefore be used to approximate the joint distribution to generate a histogram of seasonal quantitated *S. damnosum s.l.* riverine larval habitat probability error distributions, for example, to approximate the marginal distribution of one of the sampled seasonal variables, or some subset of the variables (e.g., the unknown parameters or latent variables); or to compute an integral such as the expected value of one of the variables.

Additionally, current PROC MCMC procedures provide a flexible simulation-based procedure that is suitable for fitting a wide range of Bayesianistic seasonal *S. damnosum s.l.* riverine larval habitat predictive endemic transmission-oriented risk models. For example, a vector ecologist or a local abatement district manager could, if desired, specify a likelihood function for seasonal sampled riverine larval habitat data and a prior distribution for generating a robust empirical ecological dataset of seasonal sampled riverine larval habitat parameters in PROC MCMC, while simultaneously obtaining samples from the corresponding posterior distributions. PROC MCMC will then produce summary and diagnostic statistics employing programming statements similar to those used in PROC NLMIXED. This module can then specify hyperprior distributions to fit a robust hierarchical riverine larval habitat seasonally predictive model.

In Bayesian statistics, a hyperprior is a prior distribution on a hyperparameter, that is, on a parameter of a prior distribution that arises particularly in the use of conjugate priors [2]. By saving posterior samples to an output dataset, the sampled parameters can be entered into a predictive *S. damnosum s.l.* riverine larval habitat model linearly or in any nonlinear functional form. By default, PROC MCMC uses an adaptive blocked random-walk Metropolis algorithm with a normal proposal distribution (www.sas.com). PROC MCMC can then obtain

samples from the corresponding posterior distributions, and then save the posterior samples in an output dataset which may then be used for further analysis. By so doing, the sampled *S. damnosum s.l.* riverine larval habitat data that have any likelihood, prior or hyperprior with PROC MCMC would then be accommodated as long as these functions are programmable by using the SAS DATA step functions. The seasonal-sampled parameters can then enter the model linearly or in any nonlinear functional form.

The riverine larval habitat models fit by PROC NLMIXED can also be viewed as generalizations of the random coefficient models fit by the MIXED procedure. For example, in this research, the generalization allowed the seasonal-sampled *S. damnosum s.l.* endemic transmission-oriented coefficients to enter the predictive risk model nonlinearly. By doing so, PROC NLMIXED implemented an ML estimate into the riverine larval habitat model. This is because the analogue to the REML method in PROC NLMIXED involves a high-dimensional integral over all of the fixed-effects parameters, and this integral is typically not available in a closed form. PROC NLMIXED enables users to analyze data that are normal, binomial, or Poisson by employing any likelihood programmable with SAS statements. PROC NLMIXED does not however implement the same estimation techniques available with the NLINMIX macro or the default estimation method of the GLIMMIX procedure. These are commonly based on the estimation methods of Breslow and Clayton [15], and Wolfinger and O'Connell [16], as they iteratively fit a set of generalized estimating equations. In contrast, our PROC NLMIXED constructed a predictive multivariate seasonal *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented risk model which was directly maximized employing an approximate integrated likelihood. This remark has been previously applied to the SAS/IML macros MIXNLIN [2].

Our GLIMMIX procedure also fit mixed models for non-normal data with nonlinearity in the conditional mean function. In contrast to the NLMIXED procedure, PROC GLIMMIX assumed that the model contained a linear predictor that linked the sampled endemic transmission-oriented explanatory covariate coefficients to the

conditional mean of the response. The NLMIXED procedure is designed to handle general conditional mean functions, whether they contain a linear component or not (www.sas.edu). Further, in this research, PROC NLMIXED by default performed an ML estimation by using a adaptive Gauss-Hermite quadrature. Fortunately, this estimation method is presently available with the GLIMMIX procedure (e.g., METHOD=QUAD in the PROC GLIMMIX statement).

In the future, a Bayesian inference for dynamic stochastic general equilibrium (DSGE) models should be seasonally assessed by employing a single block random walk Metropolis for quantitating georeferenced *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented explanatory covariate coefficients. DSGE models are commonly estimated by using Bayesian methods [7]. Such a predictive model may combine, adaptive independent Metropolis-Hastings and parallelization, to achieve large computational gains in a DSGE-predictive autoregressive spatiotemporal *S. damnosum s.l.* riverine larval habitat endemic transmission-oriented risk-based estimation matrix. The history of the draws would then be used to continuously improve a *t*-copula proposal distribution, and an adaptive random walk step could be inserted at predetermined intervals to escape outliers. In probability theory and statistics, a copula is a kind of distribution function used to describe the dependence between random variables [1]. A prior distribution for the seasonal predictive riverine larval habitat model parameters can then be updated to a posterior distribution using likelihood information, with sampling from the posterior carried out by using MCMC inferences.

A key feature in spatial statistical literature is the almost exclusive use of the single-block random walk Metropolis (RWM) algorithm to sample from the posterior distribution of sampled model parameter estimators. Interestingly, there is little research that tries to compare the performance of alternative sampling schemes and develop potentially better MCMC schemes' for DSGE models. Simulation efficiency is important in seasonal predictive *S. damnosum s.l.* riverine larval habitat models as MCMC inference is very time consuming for DSGE models which considerably slows down the process of model development. The

purpose of the model would be to evaluate adaptive MCMC algorithms applied to the estimation of optimal DSGE -riverine larval habitat endemic transmission oriented residual forecasts.

The main element of adaptive sampling schemes could then be the use of previous MCMC draws for the design of accurate predictive model delineating seasonal *S. damnosum s.l.* larval densities. As such, the samplers evaluated would be based on four fundamental ideas. The first is to use the history of posterior draws to repeatedly estimate *t*-copula *S. damnosum s.l.* riverine larval habitat seasonal count densities with mixtures of normal marginals, and thereafter employ these as proposal distributions in an independence Metropolis-Hastings sampler. The Metropolis-Hastings algorithm can draw samples from any probability distribution $P(x)$, provided a user (e.g., vector ecologist) can compute the value of a function $f(x)$, which is proportional to the density of P. Second, in order to alleviate some potential shortcomings of a pure independence chain approach in high-dimensional seasonal multivariate *S. damnosum s.l.* riverine larval habitat predictive risk map, simple hybrid deterministic cycling algorithms can be proposed which occasionally may use random walk proposals to escape points in the posterior parameter space especially when the posterior-to-proposal ratio is large. Third, since the time per posterior draw would not increase significantly in comparison with the other samplers due to the fast estimation of mixture of normal and *t*-copula densities, more estimators may be forecasted. Fourth, the preferred algorithms would then be suitable for parallel implementation in a robust riverine larval habitat predictive endemic transmission oriented seasonal model. Parallel computation is becoming increasingly accessible and has the potential to drastically reduce computing time in a variety of problems, and different MCMC schemes differ greatly in their suitability for parallel implementation [2].

In conclusion in this research, robust linear correlation estimates were generated from a PROC MIXED constructed regression, which identified covariates of importance associated to prolific habitats at the

Nabere study site. To fit the model in PROC NL/MIXED, the REPEATED statement was used to specify the repeated measures factor from the sampled dataset of *S. damnosum s.l.* larval habitat explanatory covariates, which then identified observations that were correlated. Non-linear estimates where then analyzed with an eigenvector spatial filtering algorithm using a positive-definite covariance matrix in SAS/GIS. By doing so, we transformed all the spatiotemporal-sampled data feature attributes containing dependence into covariates free of dependence by partitioning the original sampled riverine larval habitat data and into two synthetic variates: (1) a spatial filter variate capturing latent spatial dependency and (2) a non-spatial variate that was free of spatial dependence. The geographic distribution of the sampled habitats based on the *S. damnosum s.l.* riverine larval habitats counts exhibited PSA in all models tested: like sampled larval habitat aggregated in geospace based on spatiotemporal field-sampled count data. The coefficients from the decomposition of the sampled predictor variables were then input into a Bayesian estimation matrix. The probability density function of the inverse Wishart for the riverine larval habitat model was $\frac{|\Psi|^{\frac{\nu}{2}}}{2^{\frac{\nu p}{2}}\Gamma_p\left(\frac{\nu}{2}\right)}|X|^{-\frac{\nu+p+1}{2}}e^{-\frac{1}{2}\mathrm{tr}(\Psi X^{-1})}$, where $\boldsymbol{X}$ and Ψ are $p \times p$ positive definite matrices, and $\Gamma_p(\cdot)$ is the multivariate Gamma function. A fit of a WinBUGS hierarchical Bayesian model revealed that the adjusted covariate Hanging vegetation was statistically important to prolific sampled habitats (unadjusted improvement χ^2 was – 1.299, while the adjusted χ^2 was – 0.311).

Acknowledgement

This research was supported by a grant from the Fogerty Center of the National Institutes of Health to Thomas R. Unnasch and Robert J. Novak (project # R01TW008508).

References

[1] H. G. Gauch, Multivariate Analysis in Community Ecology, Cambridge University Press, Cambridge, England, 298 pages, Chinese Edition 1989, (1982).

[2] B. G. Jacob, R. J. Novak, L. Toe, M. S. Sanfo, A. N. Afriyie, M. A. Ibrahim, D. A. Griffith and T. R. Unnasch, Quasi-likelihood techniques in a logistic regression equation for identifying similium damnosum s.l. larval habitats intra-cluster covariates in Togo, Geospatial Information Science 15(2) (2012), 117-133.

[3] R. Matzkin, Identification in non-parametric simultaneous equations models, Ecometrica 76(5) (2008), 974-986.

[4] C. Manski, Identification of endogeneous social effects: The reflection problem, Review of Economic Studies 60(3) (1993), 531-542.

[5] S. R. Cosslett, Distribution-free maximum likelihood estimator of the binary choice, Econmetric 51(30) (1983), 765-782.

[6] A. Lewbel, Nonparametric identification of a binary random factor in cross-section data, Journal of Econometrics 163(7) (2011), 163-171.

[7] N. A. C. Cressie, Statistics for Spatial Data, Revised Edition, John Wiley & Sons, Inc., New York, 1993.

[8] J. A. Nelder and R. W. Wedderburn, Generalized linear models, Journal of the Royal Statistical Society, Series A (Royal Statistical Society) 135(3) (1972), 370-384.

[9] D. A. Griffith, Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization, Springer-Verlag, Berlin, 2003.

[10] J. Wang, W. Tang, X. Li, Y. Zhou, G. Zhao, X. Wu and Y. Liu, Factorial study on global temperature rising-impact of heat dissipation from energy consumption, Acta Scientiae Circumstantiae 26(3) (2006), 515-520.

[11] B. G. Jacob, D. A. Griffith, E. J. Muturi, E. X. Caamano, J. I. Githure and R. J. Novak, A heteroskedastic error covariance matrix estimator using a first-order conditional autoregressive Markov simulation for deriving asymptotical efficient estimates from ecological sampled Anopheles arabiensis aquatic habitat covariates, Malaria Journal 8(1) (2009), 216-225.

[12] B. G. Jacob, E. J. Muturi, E. X. Caamano, J. T. Gunter, E. Mpanga, R. Ayine, J. Okelloonen, J. Pen-Mogi Nyeko, J. I. Shililu, J. I. Githure, J. L. Regens, R. J. Novak and I. Kakoma, Hydrological modeling of geophysical parameters of arboviral and protozoan disease vectors in internally displaced people camps in Gulu, Uganda, International Journal of Health Geographics (2008), 7-11.

[13] J. G. Booth and J. P. Hobert, Standard errors of predictors in generalized linear mixed models, Journal of American Statistical Association 93 (1998), 262-272.

[14] R. W. Crosskey, Distribution Records of the Black-Flies (Diptera: Simuliidae) of Nigeria and the Southern Cameroons, With a Key for their Identification in the Pupal Stage, J. West Afr. Sci. Assoc., London. 6 (1960), 27-46.

[15] N. E. Breslow and D. G. Clayton, Approximate inference in generalized linear mixed models, Journal of the American Statistical Association 88 (1993), 9-25.

[16] R. D. Wolfinger and M. O'Connell, Generalized linear mixed models: A pseudo-likelihood approach, Journal of Statistical Computation and Simulation 48 (1993), 233-243.

■