Article

Modeling *Schistosoma japonicum* Infection under Pure Specification Bias: Impact of Environmental Drivers of Infection

Andrea L. Araujo Navas [1,*], **Frank Osei** [1], **Lydia R. Leonardo** [2], **Ricardo J. Soares Magalhães** [3,4] **and Alfred Stein** [1]

[1] Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands; f.b.osei@utwente.nl (F.O.); a.stein@utwente.nl (A.S.)

[2] Institute of Biology, College of Science, University of the Philippines Diliman, UP Diliman 1101 Quezon City, Philippines; lrleonardo@up.edu.ph

[3] UQ Spatial Epidemiology Laboratory, School of Veterinary Science, The University of Queensland, Gatton 4343 QLD, Australia; r.magalhaes@uq.edu.au

[4] Child Health and Environment Program, Child Health Research Centre, The University of Queensland, South Brisbane 4101 QLD, Australia

[*] Correspondence: a.l.araujonavas@utwente.nl; Tel.: +31-534-893-599

check for updates

Abstract: Uncertainties in spatial modeling studies of schistosomiasis (SCH) are relevant for the reliable identification of at-risk populations. Ecological fallacy occurs when ecological or group-level analyses, such as spatial aggregations at a specific administrative level, are carried out for an individual-level inference. This could lead to the unreliable identification of at-risk populations, and consequently to fallacies in the drugs' allocation strategies and their cost-effectiveness. A specific form of ecological fallacy is pure specification bias. The present research aims to quantify its effect on the parameter estimates of various environmental covariates used as drivers for SCH infection. This is done by (i) using a spatial convolution model that removes pure specification bias, (ii) estimating group and individual-level covariate regression parameters, and (iii) quantifying the difference between the parameter estimates and the predicted disease outcomes from the convolution and ecological models. We modeled the prevalence of *Schistosoma japonicum* using group-level health outcome data, and city-level environmental data as a proxy for individual-level exposure. We included environmental data such as water and vegetation indexes, distance to water bodies, day and night land surface temperature, and elevation. We estimated and compared the convolution and ecological model parameter estimates using Bayesian statistics. Covariate parameter estimates from the convolution and ecological models differed between 0.03 for the nearest distance to water bodies (NDWB), and 0.28 for the normalized difference water index (NDWI). The convolution model presented lower uncertainties in most of the parameter estimates, except for NDWB. High differences in uncertainty were found in night land surface temperature (0.23) and elevation (0.13). No significant differences were found between the predicted values and their uncertainties from both models. The proposed convolution model is able to correct for a pure specification bias by presenting less uncertain parameter estimates. It shows a good predictive performance for the mean prevalence values and for a positive number of infected people. Further research is needed to better understand the spatial extent and support of analysis to reliably explore the role of environmental variables.

Keywords: schistosomiasis; pure specification bias; uncertainty; Bayesian statistics; convolution model

1. Introduction

Schistosomiasis (SCH) is a water-borne infection caused by parasitic worms known as schistosomes. People get infected by skin penetration of the infective stage of the parasite. Three schistosomes species cause the infection: *Schistosoma mansoni*, *Schistosoma japonicum*, and *Schistosoma haematobium*. Among these, *S. japonicum* is the hardest one to control due to its zoonotic life cycle [1], which includes the infection of an amphibious snail from the species *Oncomelania hupensis quadrasi* as the intermediate host, and humans and other mammals as definitive hosts [2,3]. Schistosomiasis is a disease of public health significance [4,5] since it affects more than 252 million people worldwide [6]. This especially concerns communities in tropical and subtropical areas, where access to clean water and sanitation is limited. Schistosomiasis leads to malnutrition, which causes anemia and stunted growth in school-aged children [7,8].

Schistosomiasis risk mapping has enabled the identification of at-risk populations to target mass drug administration campaigns for disease control [9]. Mapping SCH involves the use of geographic information systems, remote sensing, and global positioning systems (GPS). These help to allocate data about infection and the physical and biological environmental variables in space. Environmental variables together with various statistical methods have been combined to model the distribution of the disease [10–13] and to quantify the role of the environmental exposure factors on SCH risk [14].

Spatial epidemiological studies are susceptible to uncertainties, which are inherent in spatial information [14–16]. Most of these uncertainties are caused by positional measurement errors due to GPS inaccuracies, multiple addresses, geocoding errors, misalignments between covariates and disease outcomes, and disease outcome or covariate aggregations [14]. Particularly, disease outcome and covariate aggregations at a specific administrative level could be incorrectly obtained for individual-level inference [17]. Health data are often available at a specific administrative level (i.e., group-level) while environmental data consists of a set of recorded values aggregated at monitor sites or gridded data derived from remote sensing images [18]. Spatial aggregation occurs due to the lack of geolocated information at the individual level, caused by the scarcity of sampling resources, availability of associated data or the need to protect confidentiality [14].

Disease outcome and covariate aggregations cause an ecological bias or ecological fallacy. This represents an important source of uncertainty [14,15,17–19] because any direct link between exposure and health outcomes is imperfectly measured [17]. For instance, Wakefield and Lyons [20] mention that the fundamental problem of this kind of spatial aggregations is the loss of information. Thus, the function used in the regression modeling does not represent the real relationship between the affected population and their exposure. This type of ecological fallacy is also called pure specification bias and arises due to the loss of information when a non-linear model changes its form under aggregation [20,21]. It is called 'pure' because it specifically addresses a model specification bias [21].

Several efforts have been made to address pure specification bias resulting in disaggregation methods. For instance, Prentice and Sheppard [22] suggest an 'aggregated data' method to create models based on exposure information available for a subset of individuals. Richardson et al. [23] assume parametric distributions for within-area exposures to derive accurate risk functions. Wakefield and Shaddick [18] propose a convolution model and derive an appropriate likelihood function for a scenario where health outcome data are aggregated at the district level, and exposure information is known at monitoring sites. Wang et al. [24] address pure specification bias in the least informative data scenario for aggregated disease counts with associated counts of the population at-risk, and a separate set of point level exposures from monitoring stations. They propose a conceptual probability of the incidence surface over the entire study region as a function of an exposure surface. This probability surface was then used to simulate individual disease outcomes and to obtain individual-level parameter estimates.

For a tropical disease such as SCH, the availability of individual-level infection data obtained from schools or health care centers is common [25–29]. Nevertheless, there are several SCH epidemiological studies [9,30–33] that limit their modeling to the outcome and covariate data aggregated at a specific

administrative level (i.e., ward, municipal, province, county, district, barangays, among others) without taking pure specification bias into consideration. Moreover, there are no up-to-date studies that have quantified the effect of the covariate information on the parameter estimates when accounting for pure specification bias in SCH modeling.

The objective of this study is to quantify the effect of pure specification bias on the parameter estimates of various environmental covariates used as drivers for SCH infection. To achieve this objective we aim to: (i) use a spatial convolution model that removes pure specification bias by using group-level health outcome data and individual-level environmental covariate data, (ii) estimate group and individual-level covariate regression parameters, and (iii) quantify the difference between the parameter estimates and predicted disease outcomes from the convolution and ecological models.

2. Materials and Methods

2.1. Data on Human Schistosoma japonicum *Infection, Study Area, and Sampling Design*

We use *S. japonicum* infection data collected as part of the 2008 Nationwide Schistosomiasis Survey in The Philippines [34]. In this case, *S. japonicum* is endemic in 28 of its 81 provinces [35], with approximately 1.8 million estimated infected people [36]. The disease affects children, adolescents, and individuals with high-risk occupations, such as farmers and fishermen [36,37].

A two-stage systematic cluster sampling was used in the sampling design. Stratification was done by a prevalence level (high, medium, and low), obtained from the 1994 World Bank-assisted Philippine Health Development Program. Provinces and barangays were the primary and secondary sampling units respectively. A barangay is the smallest administrative division in the Philippines, numbering about 22–50 in a single municipality. Provinces with high and moderate prevalence rates were included, while the random selection was done among the low-prevalence provinces and non-endemic provinces. Within the selected provinces, high prevalence barangays were also selected. We decided to work in the Mindanao region due to the good spatial coverage of the sampling over the whole region (Figure 1), and the high response rate of 70 percent [9,34,38]. In total, 22 provinces were surveyed, and between 2 and 10 barangays were surveyed per province. In total, 108 out of 10,021 barangays were surveyed in Mindanao.

Data from 19,763 individuals were recorded in the survey but not georeferenced. Information regarding the corresponding barangay and province were recorded for each individual. For this reason, individual-level survey data were aggregated and geo-located to the centroids of 108 barangays in Mindanao. We used a probability of infection p in barangay k as our disease outcome variable.

Kato-Katz thick smear examination [39] was used to diagnose *S. japonicum* infection based on two-sample stool collection. Each sample was read using a microscope and the presence of *S. japonicum* eggs indicated active infection. Due to inconsistencies in the submission of the second stool sample, however, only the results of the first stool sample were available [9] from people aged two years and above. Information such as gender and age were recorded for each individual. More details about the sampling design can be found in Leonardo et al. [34,38].
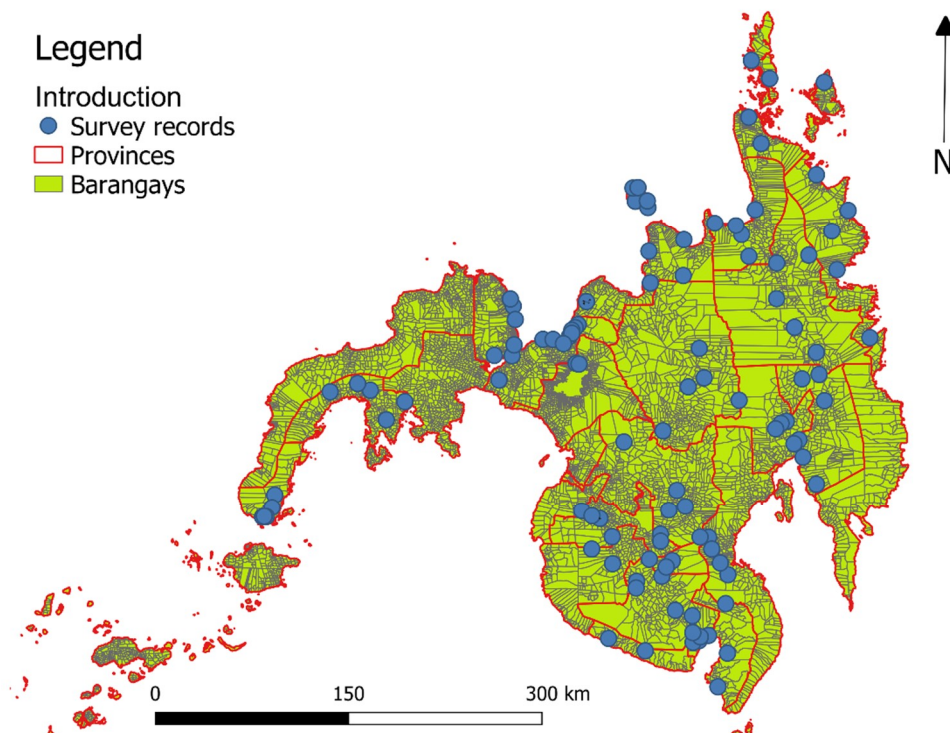
Figure 1. Study area.

2.2. Environmental and Geographical Data

Six environmental variables were included in the analysis: the nearest distance to water bodies (NDWB), normalized difference vegetation index (NDVI), normalized difference water index (NDWI), day (LSTD) and night (LSTN) land surface temperature, and elevation (E). The nearest distance to water bodies was calculated using the closest facility network analysis tool in ArcGIS version 10 [40]. We used rivers and lakes as water bodies. As input for the network, we used the river and road networks, and the location of cities and hamlets. We first calculated the NDWB for each city point and then interpolated the distance values for all the surveyed barangays. Interpolation was performed using Ordinary Kriging. The nearest distance to water bodies shows the accessibility of people to water bodies since they represent the main infection foci. We used NDVI obtained from the Modis MOD13Q1 product. The normalized difference vegetation index served as an indicator of vegetation presence and greenness, particularly the presence of flooded agricultural land such as paddy fields, which is an important factor for Asian schistosomiasis [41]. We included NDWI from the Google Earth Engine as an annual Landsat 7 composite for the year 2008. The normalized difference water index was used as a proxy indicator of flooding [42,43]. The day land surface temperature was also included and obtained from the Modis MOD11A2_LST product. The day land surface temperature is determinant for the survival of larval stages of snails [44–46] and is used as a proxy for water temperature given that the thermal condition of shallow waters usually reflects the ambient temperature of the air [9]. Elevation was also included and obtained from ASTER GDEM version 2 from United States Geological Survey (USGS) [47]. In the Philippines, the presence of snails is also driven by the local topography [48–50]. At lower altitudes, the risk of finding snails increases. Table 1 summarizes the information about environmental information and their sources.

Table 1. Environmental variables description.

Environmental Variable	Spatial Resolution	Temporal Resolution	Data Type	Original Coordinate System	Data Source
Elevation	30 m	NA	Raster	EPSG:4326	Aster GDEM V2 from USGS
NDVI	250 m	2008	Raster	EPSG:4326	MOD13Q1
NDWI	500 m	2008	Raster	EPSG:32651	Landsat 7, one-year composite
LST	1 km	2008	Raster	EPSG:4326	MOD11A2
NDWB	250 m	2010	Raster	EPSG:32651	Derived from closest facility network using roads, urban areas, river network, and water bodies

NDVI: Normalized difference vegetation index; NDWI: normalized difference water index; LST: day land surface temperature; NDWB: nearest distance to water bodies; USGS: United States Geological Survey.

2.3. Convolution Model (Individual-Level Model)

An individual is considered infected if at least one parasite egg is found. *S. japonicum* infection data y are available at individual-level i recorded within a barangay k. Various n environmental variables x are available for each image pixel. Because the exact response locations of the y_{ik} are unknown, individual-level data are aggregated to their corresponding barangay centroid and are, thus, denoted by y_k. A naive group-level model is given by Equation (1).

$$y_k | \overline{x}_k, \gamma \sim Binomial(N_k, \hat{p}_k) \tag{1}$$

where N_k and p_k are the number of sampled individuals and the probability of infection in barangay k, respectively, and $\overline{x}_k$ is the observed mean exposure within barangay k. The probability $\hat{p}_k$ is modelled based on Equation (2) using environmental variables as predictors, where γ are the group-level covariate coefficients.

$$logit(\hat{p}_k) = \gamma_0 + \gamma_1 \cdot \overline{x}_{1k} + \gamma_2 \cdot \overline{x}_{2k} + \ldots + \gamma_n \cdot \overline{x}_{nk} \tag{2}$$

We suppose that, for an individual i in area k, y_{ik} follows a Bernoulli distribution (Equation (3)). Then an individual level model is presented in Equation (4), where the parameters β are the individual-level coefficients. However, this model assumes that we know the individual level locations.

$$y_{ik} | x_{ik}, \beta \sim Bernoulli(p_{ik}) \tag{3}$$

$$logit(p_{ik}) = \beta_0 + \beta_1 \cdot x_{1ik} + \beta_2 \cdot x_{2ik} + \ldots + \beta_n \cdot x_{nik} \tag{4}$$

Pure specification bias will result in $\gamma \neq \beta$, where the relationship between aggregated disease risk and exposure on areal units differs from the relationship between the individual disease risk and the associated exposure. In a non-linear model, as in our case, this difference is produced by a loss of information due to aggregation known as pure specification bias. Pure specification bias is reduced in size since the 'within area' exposure is more homogenous [20]. This could be obtained by having a finer partition of space in which exposure measurements are available [18,20].

As we know the individual-level responses y_{ik} but not their locations, we minimized the pure specification bias by extracting covariate information from cities within the barangays. We selected cities since they were the finest units we found available by taking them as a proxy for exposure locations at an individual level. Cities were extracted from the 2010 build-up data base from the National Mapping and Resource Information Authority from The Philippines [51]. Cities were not available for all the surveyed barangays. Therefore, we digitalized them using Google Earth images.

We used the aggregate data method proposed by Prentice and Sheppard, 2001 [22]. Let exposure or covariate data x_{jk} be measured at locations s_{jk} $j = 1, \ldots, m_k \leq N_k$, for a subset of individuals. Then, we estimated the average risk of the individuals in area k, and the individual level coefficients β. This

was done by calculating the mean of the risk function (Equation (5)) instead of evaluating the risk function at the mean exposure (Equation (2)). In this way, the average $\hat{p}_k$ of the function over the exposures corrects for the pure specification bias and differs from the function evaluated at the average exposure ($\hat{p}_k$). Thus, $\hat{p}_k$ is the estimated average probability of infection of the individuals in area k.

$$\hat{p}_k = \frac{1}{m_k} \cdot \sum_{j=1}^{m_k} \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 \cdot \overline{x}_{1jk} + \beta_2 \cdot \overline{x}_{2jk} + \ldots + \beta_n \cdot \overline{x}_{njk}\right)\right)} \tag{5}$$

In our study, covariate data were obtained at a finer level of analysis than the barangay. Therefore, we assumed that averaged covariate values at city level j represent individual covariate values (i.e., $x_{jk} = x_{ik}$). We then know x_{jk} but not the geographical linkage with individuals. One way to account for this is to allocate N_{jk} individuals to measurement x_{jk} by equally dividing the population. Therefore, $N_{jk} = N_k/m_k$, is a simple version of the convolution model.

The spatial convolution model is represented in Equation (6), where the risk function also accounts for the spatial variability by including spatial (s_k) structured random effects.

$$\hat{p}_k = \frac{1}{m_k} \cdot \sum_{j=1}^{m_k} \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 \cdot \overline{x}_{1jk} + \beta_2 \cdot \overline{x}_{2jk} + \ldots + \beta_n \cdot \overline{x}_{njk} + s_k\right)\right)} \tag{6}$$

The implemented convolution model that corrects for the ecological fallacy is of the form.

$$y_k | \overline{\boldsymbol{x}}_{jk}, \boldsymbol{\beta} \sim Binomial\left(N_k, \hat{p}_k\right) \tag{7}$$

The model includes an intercept (β_0), averaged city-level environmental variables ($\overline{x}_{jk} = NDVI, NDWI, LSTD, LSTN, E, NDWB$), and their corresponding individual-level coefficients $\boldsymbol{\beta}$, and a spatial random effect (s_k) as described in Equation (6). All covariates were standardized to have mean = 0 and standard deviation = 1. Collinearity between covariates was assessed with the Pearson correlation coefficient.

Prior information for the intercept β_0 was given as a proper uniform distribution with wide bounds ($-100, 100$). The other $\boldsymbol{\beta}$ parameters were given non-informative normal distributions with mean = 0 and precision $1/\sigma^2$, with σ uniformly distributed on a wide range $(0, 100)$. These distributions are recommended in order to avoid overestimations on the parameters [52]. These parametrizations allow a good mixing of the sequences used for Marcov Chain Monte Carlo simulations and contribute to their faster convergence [53].

Spatial dependence was modelled using a spatially structured random effects distribution based upon a geo-statistical model. This model can be used as a sampling distribution for continuous spatial data [54]. The vector of random variables $\boldsymbol{s}$ associated with point locations (x_k, y_k), $k = 1, \ldots, K$, was modelled with a multivariate normal distribution $s \sim MVN_K(\mu, \Sigma_{ab})$ with mean $\mu = 0$ and a covariance matrix Σ_{ab} defined by a powered exponential spatial correlation function from Equation (8).

$$\Sigma_{ab} = \sigma^2 \cdot \exp\left[-(\phi \cdot d_{ab})^\kappa\right] \tag{8}$$

The covariance matrix is specified as a function of the distances d_{ab} between barangay centroids a and b, the rate of decline of spatial correlation per unit of distance ϕ, the scalar parameter representing the overall variance σ^2, and the scalar parameter κ controlling the amount of spatial smoothing. Since it is often difficult to learn much about the κ parameter, and large values of κ could lead to smoothing, we used $\kappa = 1$. The prior distribution for ϕ was set uniform with lower and upper bounds at 2×10^{-7} and 3×10^{-3}, respectively. These values give a diffuse but plausible prior range of correlations between 0.1 and 0.99 at the minimum distance between points (575 m) and between 0 and 0.3 at the maximum distance between points (<552 km), which assists identification [55]. The variance parameter was

given a half-normal distribution with mean = 0 and variance = 1. Half-normal was selected in order to restrict our prior to positive values and avoid problems with convergence [52,56].

The model was run using three sequences or chains with 50,000 iterations that ensured simulations representative of target distributions [53] and a good stability for convergence [53]. A burn-in of 25,000 iterations was used, discarding the first half of each sequence that is used to diminish the influence of starting values [53]. We monitored convergence visually and statistically, first by inspecting at the trace plots and then by checking the $\hat{R}$ statistic [57,58], which are also called a potential scale reduction factor. This assesses sequences mixing by comparing the between and within variation. $\hat{R}$ values < 1.1 indicate evidence that sequences had converged [57], while high values suggest that an increase in the number of simulations may improve our inferences [53]. Data were structured in a rectangular format, where the columns are headed by the array name. The survey data and the codes in bugs for the convolution and ecological models are provided in the supplementary files 1 and 2, respectively.

2.4. Ecological Model (Group-Level Model)

As a comparison, we estimated the group-level covariate regression parameters γ by using Equations (1) and (2). We used $\hat{p}_k$ as a function of the same covariate information, $\overline{x}_k = NDVI, NDWI, LSTD, LSTN, E, NDWB$, but averaged for each surveyed barangay, and added the spatial random effects term s_k (Equation (9)). Prior distributions for γ and s_k were the same as the ones given for the β parameters and the geo-statistical spatial term.

$$\hat{p}_k = \sum_{k=1}^{K} \frac{1}{1 + \exp(-(\gamma_0 + \gamma_1 \cdot \overline{x}_{1k} + \gamma_2 \cdot \overline{x}_{2k} + \ldots + \gamma_n \cdot \overline{x}_{nk} + s_k))} \tag{9}$$

We compared estimated covariate regression coefficients and their credible intervals from both models, and also the data generated from the posterior predictive distribution to the observed data. Posterior predictive distributions were generated using simulations. Residuals were calculated by subtracting the simulated values from the observed values. We created correlation and residual plots for convolution and ecological models for three different simulations. Lastly, we compared the predicted prevalence values for both models and their corresponding credible intervals for each barangay.

2.5. Model Validation

In order to assess the model fit, we compared the deviance information criterion (DIC) values between simple and spatial models from the convolution and ecological models. Convolution and ecological models were validated using two methods. First, we used the posterior predictive distribution to check our model assumptions by comparing the data generated from the simulations of the predictive distribution to the observed data using a test statistic. The test statistic generates a posterior predictive *p*-value (pp*p*-value) by calculating the proportion of the predicted values that are more extreme for the test statistic than the observed value for that statistic. If the data violated our model assumptions, the observed test statistic should differ from most of the replicated test statistics from our model (i.e., pp*p*-value close to 0 or 1). If the model fits the data, a pp*p*-value around 0.5 is expected. Test statistics and pp*p*-values were created for maximum, minimum, and mean values for both, convolution, and ecological models.

Second, we used the area under the curve (AUC) of the receiving operating characteristics (ROC) for applying a threshold of 0.5%, which is the prevalence mean in the Mindanao region. Thus, we would like to know the ability of the model to discriminate the mean prevalence level in the study area. We also investigated the ability of the models to discriminate the number of positive cases. Thus, a threshold of 1 was used, which indicates the presence of at least one positive case. An AUC value of 70% was taken to indicate acceptable predictive performance [9,59].

2.6. Software and Data Sources

Barangay centroids were obtained from an up-to-date barangay shape file from a DIVA geographic information system [60]. River and road networks were obtained from the Open Street Map Project in the Philippines [61]. Locations of cities and hamlets were extracted from the National Mapping and Resource Information Authority from The Philippines [62] from 2010.

The model was implemented in the software OpenBUGS 3.2.3 [63,64] (Medical Research Council, Cambridge, UK and Imperial College London, UK). The software is available for free at [65]. We used the package R2OpenBUGS [66] to call OpenBUGS from R. The spatial models were coded using functions from GeoBUGS [55], which is an add-on module to OpenBUGS that provides an interface to work with conditional autoregressive and geo-statistical models. Data pre-processing and Ordinary Kriging was performed in R [67].

2.7. Ethics Approval

The data used in this study was collected in 2005 when there was no requirement for ethical review and clearance. This study used aggregated (barangay-level) survey data, which enabled full de-identification of individuals involved in the survey.

3. Results

3.1. Convolution Model

Posterior means and credible intervals resulting from the simple version of the convolution model (Equation (5)) are given in Table 2. The credible intervals did not include zero values, which shows that all covariates have a strong effect in the observed outcomes except NDVI. We decided to discard NDVI from the spatial convolution and ecological model (Equation (6)). Deviance information criterion values for the simple and spatial models were 419.8 and 64.13, respectively. This shows that the spatial model performs better than the simple model. This is also supported by the residual spatial variation of the survey data presented in Mindanao [9]. Table 2 shows the resulting parameter estimates and credible intervals for the spatial convolution model.

Table 2. Estimated regression coefficients (mean and 95% credible intervals).

Estimated Parameters	Posterior Mean (95% Crl)		Standard Deviation		Credible Intervals Width (Uncertainty)	
	Convolution Model	Ecological Model	Convolution Model	Ecological Model	Convolution Model	Ecological Model
Intercept	$-5.79 \, (-6.11, -5.5)$	$-5.67 \, (-5.93, -5.41)$	0.16	0.13	0.62	0.53
NDWI	$-0.74 \, (-0.96, -0.55)$	$-1.02 \, (-1.24, -0.80)$	0.11	0.11	0.44	0.44
LSTD	$-0.63 \, (-0.92, -0.38)$	$-0.79 \, (-1.08, -0.49)$	0.14	0.15	0.56	0.59
LSTN	$-0.84 \, (-1.13, -0.55)$	$-0.65 \, (-1.05, -0.24)$	0.15	0.21	0.59	0.82
Elevation	$-1.05 \, (-1.4, -0.71)$	$-1.13 \, (-1.53, -0.70)$	0.18	0.22	0.71	0.84
NDWB	$-0.28 \, (-0.51, -0.05)$	$-0.24 \, (-0.43, -0.05)$	0.13	0.09	0.48	0.38
ϕ	$4 \times 10^{-5} \, (-0.004, 0.004)$	$2 \times 10^{-5} \, (-0.0004, 0.0004)$	2.00×10^{-4}	1.00×10^{-5}	6.80×10^{-5}	3.50×10^{-5}
Variance of spatial random effect	2.58 (1.7,3.6)	2.6 (1.8,3.61)	0.48	0.47	1.9	1.82

Crl: Credible interval.

3.2. Ecological Model

As in the convolution model, credible intervals resulting from the simple version of the ecological model (Equation (2)) showed that all covariates have a strong effect in the observed outcomes except NDVI. The normalized difference vegetation index was also discarded from the spatial ecological model (Equation (9)). Deviance information criterion values for the simple and spatial models were 388.1 and 126.4, respectively. This shows that the spatial model is more adequate. Table 2 shows the resulting parameter estimates and credible intervals for the spatial ecological model.

3.3. Convolution Versus Ecological Model

Figure 2 shows the resulting density plots from the regression parameter estimates derived from the ecological and convolution models. For the intercept parameter (Figure 2a), the convolution model estimated a lower mean value (difference = 0.11) (Table 2) with a higher uncertainty, or credible interval width, than the ecological model (Table 2). The regression parameter estimate for NDWI from the convolution model shows a higher mean value than the one from the ecological model (Table 2 and Figure 2b). The difference between these estimates is around 0.28 with the same uncertainty for both models (Table 2). In the case of LSTD (Figure 2c), the estimated mean values from the convolution model are higher but with a slightly lower uncertainty than the ecological model. Difference between estimates are around 0.16 (Table 2). For the LSTN parameter (Table 2) and Figure 2d, the convolution model estimated a lower mean value with lower uncertainty than the ecological model (Table 2). The difference between these estimates is approximately 0.2. In the case of the elevation parameter (Figure 2e), estimated mean values from the convolution model are slightly higher than estimates from the ecological model (difference = 0.08) (Table 2) and have lower uncertainty than the ecological model estimates (Table 2). The estimated parameter value for NDWB (Figure 2f) is lower in the convolution than in the ecological model (Table 2). The difference in this value is relatively small at around 0.03. Nevertheless, uncertainty is higher in the convolution than in the ecological model (Table 2).
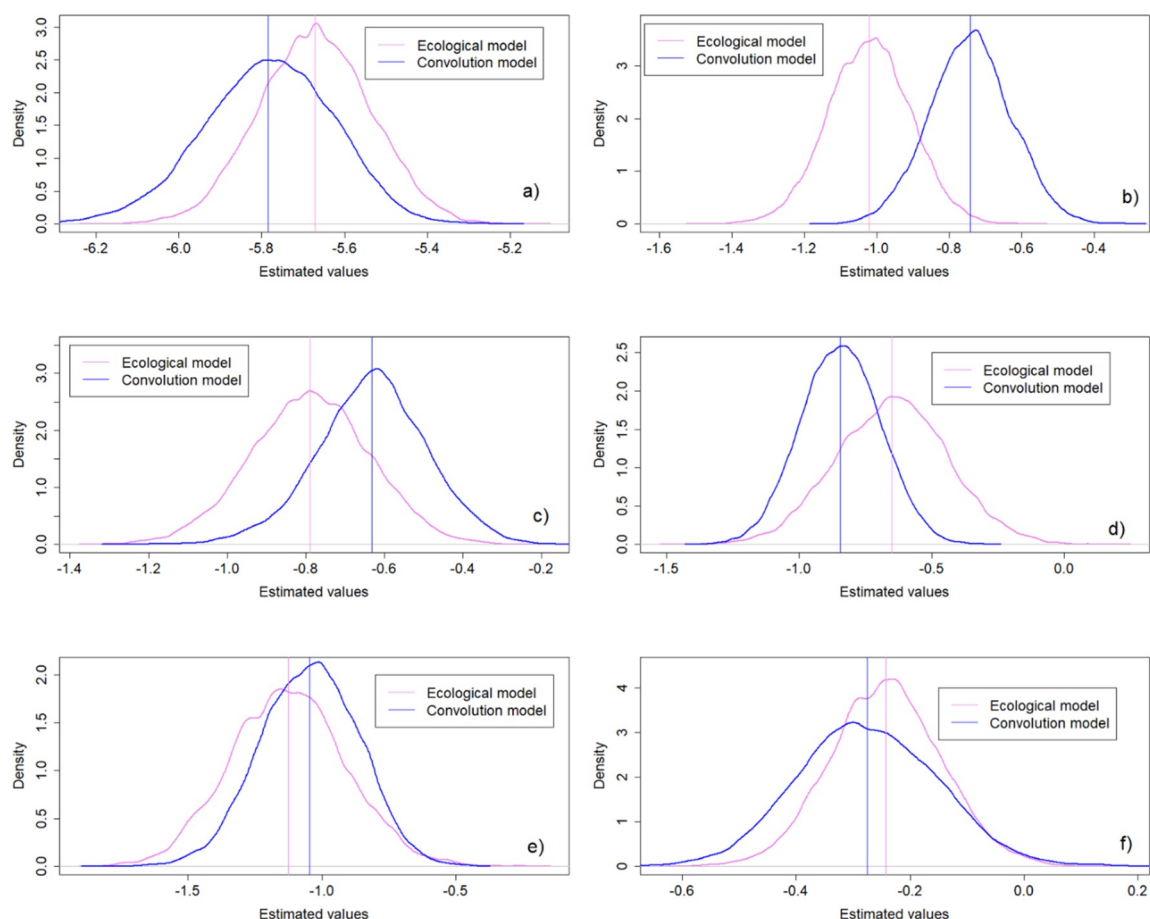

Figure 2. Covariate regression coefficients density plots for the convolution and ecological models. (**a**) Intercept, (**b**) Normalized Difference Water Index, (**c**) Day Land Surface Temperature, (**d**) Night Land Surface Temperature, (**e**) Elevation, and (**f**) Nearest Distance to Water Bodies.

Differences between observed and predicted prevalence values are similar for both models ($R^2 = 0.9$). For the convolution and ecological model, the maximum difference between the predicted and the observed prevalence values is around 1% (Figure 3a,b). Figure 3a,b show that, for fitted

prevalence values higher than 2%, both models underestimate the prevalence of infection, while, for fitted prevalence values lower than 2%, both overestimation and underestimation occur. Both models show similar predicted values with a maximum difference of 0.3%. The uncertainty in the predictions from both models is the same in 88 barangays. The ecological model presents a higher uncertainty than the convolution model in 9 barangays, while the convolution model shows higher uncertainty than the ecological model in 11 barangays.
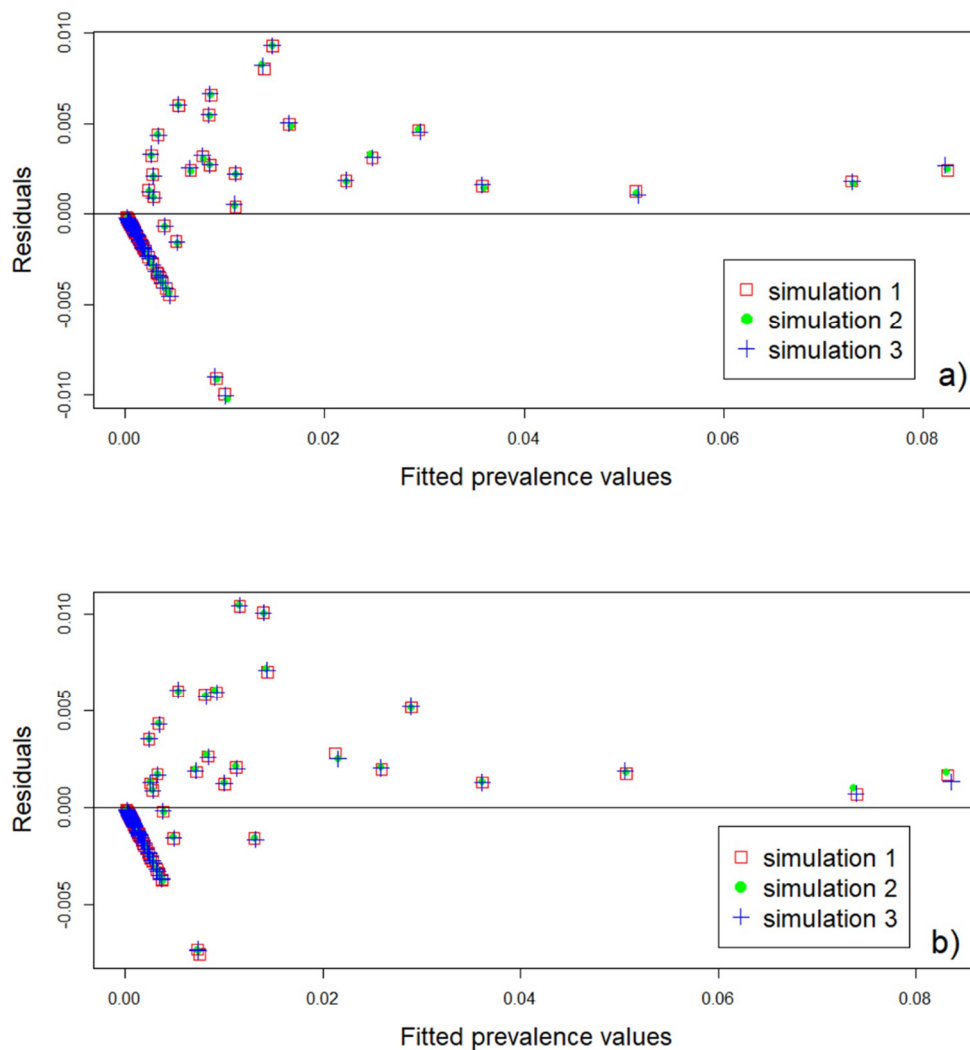
Figure 3. Residual plots for the (**a**) convolution and (**b**) ecological models.

3.4. Model Validation

The maximum and minimum observed prevalence values are 0.085 and 0, respectively. For the convolution model, the pp$p$-values for the maximum and minimum observed values are 0.65 and 1, respectively. This shows that it is likely to see the maximum and minimum prevalence values from the observed data in the predicted data. The highest pp$p$-value of 1 assures that 100% of our predicted data contain the minimum observed value. This could be due to an over fit to the data for small prevalence values. For the ecological model, the maximum and minimum pp$p$-values are 0.67 and 1, respectively. Like the convolution model, it is likely to see the maximum and minimum observed prevalence values in our predicted data and there might be an over fit to the data for small prevalence values.

Results from the second validation method show the high ability of both models to predict prevalence values with an AUC equal to 93% and 94% for the convolution and ecological models, respectively. The AUC values with respect to predictive ability of the number of cases are 81% and

94% for the convolution and ecological models, respectively. Both models can discriminate the number of positive cases of schistosomiasis and mean prevalence values. Nevertheless, as we can see, the ecological model might over fit the data as compared to the convolution model with respect to the number of positive cases.

4. Discussion

Several studies have modelled SCH disease risk using surveyed and environmental aggregated information at an administrative-level [9,30–33]. These studies so far ignored the pure specification bias caused by the use of ecological or group-level estimates as individual-level estimates. Only a few studies [18,24] have quantified the influence of pure specification bias on the regression parameter estimates and all studies on SCH ignored the influence of pure specification bias on disease predictions. In our paper, we quantified the effect of pure specification bias on assessing the parameters for environmental covariates that are used for the mapping of *S. japonicum* infection risk. Our contribution is both methodological and practical.

Our starting point was that NDVI, NDWI, LSTD, LSTN, elevation, and NDWB are relevant for SCH transmission [68–70]. For instance, NDVI is an indicator of flooded vegetation [9], specifically rice paddy fields, and environmental moisture [42,71]. In both models, all variables have a strong effect in the observed range of outcomes, except NDVI. An explanation could be the effect of spatial support of this variable at 250 m. The International Rice Research Institute (IRRI) estimated an area substantially smaller than 25 ha for rice paddy fields [72]. The area covered by an NDVI pixel equals 6.25 ha. This shows that a spatial support of 250 m is still too coarse to reliably represent paddy fields. It could be relevant to use a higher spatial support to reliably assess the role of NDVI. For instance, Walz et al. [42] have successfully delineated paddy fields by using NDVI from RapidEye imagery at a higher spatial support of 5 m.

In the case of NDWI, the convolution model estimates a higher NDWI parameter value than the ecological model, but closer to zero (Figure 2b). Hence, it may not have a strong effect on the observed range of outcomes. The difference between convolution and ecological models when estimating the NDWI parameter is high (0.28) as compared to the differences for other variables. This could be due to (i) the decrease of the spatial extent of analysis from barangay to a city-level for covariate extraction, and (ii) the coarse spatial support of the variable at 500 m. A correspondence between a decrease in the extent of analysis and the within cities variability of NDWI values would change the average value used for parameter estimation. This could yield the high differences in the estimates and, together with the coarse support of the variable, could lead to a weaker effect of NDWI in SCH prevalence. Hence, NDWI pixels of 0.25 km^2 are too coarse to reliably define flooded zones in city areas that range from 0.02 to 3 km^2. The uncertainty in NDWI parameter estimate is similar for both models, as shown by the credible interval with of 0.44 (Figure 2b and Table 2). A possible explanation is that NDWI values are similar between and within cities, and between and within barangays. For instance, NDWI values between cities and between barangays range from 0.095 to 0.51 and from 0.099–0.51, respectively. The average NDWI value within a specific barangay is 0.3, while NDVI values range from 0.29–0.31 for the cities it contains.

For LSTD, a higher estimated value in the convolution model was observed than in the ecological model (Figure 2c). As for NDWI, this could be (i) due to the decrease in the within cities variability corresponding to a decrease in spatial extent for covariate values extraction, and (ii) the coarse LSTD spatial support of 1 km. The day land surface temperature, area pixels of 1 km^2 are too coarse to reliably define low and high-temperature zones in city areas ranging from 0.02 to 3 km^2. The uncertainty is slightly lower in the convolution model than in the ecological model possibly due to the similarity of the LSTD variability between and within cities and between and within barangays. For instance, similar LSTD values ranging from 25 to 35.7 °C and from 21.7 to 36 °C were found for the convolution and ecological models, respectively.

A different pattern is observed for the estimate of the LSTN parameter. In contrast to LSTD, the estimate and uncertainty from the convolution model are lower than for the ecological one (Figure 2d). This could be explained by the difference between the LSTN variability within cities and within barangays. For instance, we selected a barangay bigger in size than the cities it contains. We compared LSTN values from the cities and from the barangay and found differences from 4 to 7 °C, although the spatial support of LSTN is coarse at 1 km. These differences are small, but could be determinant for the parasite presence, as the distribution of SCH is driven by water temperatures from 15 to 20 °C [73]. This means that we could find the parasite in the barangay with an average LSTN value of 20 °C, but we could not find it in the cities inside the barangay, with LSTN values ranging from 22–24 °C. Thus, there is a clear loss of information produced by a pure specification bias.

Elevation presents an estimate closer to zero in the convolution model as compared to the ecological model (Figure 2e). This is possibly related to the decrease in the within cities elevation variability as a result of the decrease in the spatial extent of covariate extraction, from barangays to cities. Although its spatial support is relatively high, i.e., 30 m, within cities variability decreases because the Mindanao region does not contain steep slopes or sharp changes in elevation. Changes in elevation are gradual, which means that a city can share a single elevation value. The uncertainty in the convolution model is lower than in the ecological model. The reason could be the large difference in the between and within cities and barangays variability. For instance, elevation values between the cities ranged from 0 to 0.9 km, while elevation values between barangays ranged from 0 to 1.3 km. We also compared the elevation values averaged in a specific barangay and the range of values from the cities within the barangay. The averaged barangay value was around 1.3 km, while the city values within the barangay ranged from 0.87 to 0.89 km. These differences are high and give an idea of the large amount of information that could be lost when estimating individual parameters at ecological levels of analysis.

Lastly, for NDWB, the convolution model estimates a lower parameter value but with a higher uncertainty (Figure 2f). This could be explained by the discrepancy between the within barangay and city variability. For instance, for a specific barangay, the averaged NDWB value was approximately 3.75 km, while NDWB values for the cities within this barangay ranged from 0.4 to 6.4 km. In addition, the NDWB values between the cities ranged from 0.17 to 26.2 km, while the NDWB values between the barangays ranged from 0.26 to 15.5 km. The higher uncertainty in the convolution model could be explained by the use of ordinary Kriging in the NDWB calculation. The use of interpolation increases the variance in the estimates in a somewhat unrealistic way since it uses a constant mean [54] while, in reality, the different cities and barangays have different means.

The present research shows that the ecological and convolution models present similar prediction results. However, the proposed convolution model is preferred based on the lower uncertainties found in most of its parameter estimates, which shows that it corrects for pure specification bias. Moreover, according to our validation results, the convolution model has a high predictive ability to detect a positive number of cases (81%) and mean prevalence values (93%). From a public health perspective, the provision of regression coefficients that are less uncertain and better approach the individual-level estimates is a step forward to the desired uncertainty analysis in a schistosomiasis-modelling framework. This could be used as a decision support tool for helminth control programs. Moreover, less uncertain models and maps would avoid erroneous conclusions and decisions about the spatial distribution of schistosomiasis. Lastly, information on uncertainty regarding pure specification bias could guide mass drug administration campaigns by enhancing the assessment of the infection risk and understand potential impacts on human health [15,74].

Suggestions

Spatial extent and support of analysis are relevant drivers for the model parameter estimates and their associated uncertainties. The choice of support may affect the pattern identified from the data and the relationship between environmental variables and SCH prevalence [75,76]. We recommend

bringing all covariates to a common spatial support prior to analysis. A suggestion would be to start at e.g., 30 m and examine larger supports to more precisely quantify the role of an environmental variable in the disease modelling process [13,77].

The trend in the residual plots from both models (Figure 3) points to a dependence between the residuals and the fitted prevalence values. This dependence could be due to heteroscedasticity [78], which means that similar interactions between the variables could lead to different prevalence values. This does not represent a problem in the model parameter estimates but is an indicator that the model can be improved [79,80]. Although our aim is not about model fit, we could possibly improve our predictions by exploring other disease driven factors and include them in the model or explore the model specification. Perhaps fitting a zero-inflated binomial (ZIB) model could help improve the predictions given that our data show a high number of barangays with zero prevalence (~77%).

5. Conclusions

The present study proposes a convolution model that removes pure specification bias by using ecological, group-level, health outcome data and city-level, individual-level, environmental data. For most covariates, the uncertainties in the convolution model are lower than those in the ecological model.

The spatial extent of the covariates values and the spatial support or resolution of these covariates are relevant for the parameter estimates and their uncertainties. The spatial extent and support also influence the role of the covariates in SCH modelling. Why this happens should be further explored. Additionally, between-covariate and within-covariate variability resulted in similar uncertainties in both models. Conversely, differences in between and within covariate variability explain the loss of information produced by a pure specification bias, which leads to lower uncertainties in the convolution model.

Lastly, this study shows no significant differences in the predicted values from both models. Predicted values from the convolution model are as uncertain as the predicted values from the ecological model in the majority of surveyed barangays (81.5%). The convolution model, however, shows a good predictive performance for the mean prevalence values and a positive number of infected people.

Author Contributions: Conceptualization, A.L.A.N., F.O., and R.J.S.M. Formal analysis, A.L.A.N. Investigation, A.L.A.N. Methodology, A.L.A.N. and F.O. Resources, L.R.L. and R.J.S.M. Software, A.L.A.N. Supervision, F.O., R.J.S.M., and A.S. Validation, A.L.A.N. Visualization, A.L.A.N. Writing—original draft, A.L.A.N. Writing—review & editing, A.L.A.N., F.O., L.R.L., R.J.S.M., and A.S.

References

1. Jia, T.W.; Zhou, X.N.; Wang, X.H.; Utzinger, J.; Steinmann, P.; Wu, X.H. Assessment of the age-specific disability weight of chronic schistosomiasis japonica. *Bull. World Health Organ.* **2007**, *85*, 458–465. [CrossRef] [PubMed]

2. Tarafder, M.R.; Balolong, E.; Carabin, H.; Belisle, P.; Tallo, V.; Joseph, L.; Alday, P.; Gonzales, R.O.; Riley, S.; Olveda, R.; et al. A cross-sectional study of the prevalence of intensity of infection with Schistosoma japonicum in 50 irrigated and rain-fed villages in Samar Province, the Philippines. *BMC Public Health* **2006**, *6*, 10. [CrossRef] [PubMed]

3. Yang, K.; Wang, X.H.; Yang, G.J.; Wu, X.H.; Qi, Y.L.; Li, H.J.; Zhou, X.N. An integrated approach to identify distribution of Oncomelania hupensis, the intermediate host of Schistosoma japonicum, in a mountainous region in China. *Int. J. Parasitol.* **2008**, *38*, 1007–1016. [CrossRef] [PubMed]

4. King, C.H.; Dickman, K.; Tisch, D.J. Reassessment of the cost of chronic helmintic infection: A meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* **2005**, *365*, 1561–1569. [CrossRef]

Int. J. Environ. Res. Public Health **2019**, *16*, 176

14 of 17

5. Walz, Y.; Wegmann, M.; Dech, S.; Vounatsou, P.; Poda, J.-N.; N'Goran, E.K.; Utzinger, J.; Raso, G. Modeling and Validation of Environmental Suitability for Schistosomiasis Transmission Using Remote Sensing. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0004217. [CrossRef] [PubMed]

6. Hotez, P.J.; Alvarado, M.; Basanez, M.G.; Bolliger, I.; Bourne, R.; Boussinesq, M.; Brooker, S.J.; Brown, A.S.; Buckle, G.; Budke, C.M.; et al. The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.* **2014**, *8*, e2865. [CrossRef] [PubMed]

7. Leenstra, T.; Acosta, L.P.; Langdon, G.C.; Manalo, D.L.; Su, L.; Olveda, R.M.; McGarvey, S.T.; Kurtis, J.D.; Friedman, J.F. Schistosomiasis japonica, anemia, and iron status in children, adolescents, and young adults in Leyte, Philippines. *Am. J. Clin. Nutr.* **2006**, *83*, 371–379. [CrossRef]

8. Coutinho, H.M.; McGarvey, S.T.; Acosta, L.P.; Manalo, D.L.; Langdon, G.C.; Leenstra, T.; Kanzaria, H.K.; Solomon, J.; Wu, H.W.; Olveda, R.M.; et al. Nutritional status and serum cytokine profiles in children, adolescents, and young adults with Schistosoma japonicum-associated hepatic fibrosis, in Leyte, Philippines. *J. Infect. Dis.* **2005**, *192*, 528–536. [CrossRef]

9. Soares Magalhães, R.J.; Salamat, M.S.; Leonardo, L.; Gray, D.J.; Carabin, H.; Halton, K.; McManus, D.P.; Williams, G.M.; Rivera, P.; Saniel, O.; et al. Geographical distribution of human Schistosoma japonicum infection in The Philippines: Tools to support disease control and further elimination. *Int. J. Parasitol.* **2014**, *44*, 977–984. [CrossRef]

10. Herbreteau, V.; Salem, G.; Souris, M.; Hugot, J.-P.; Gonzalez, J.-P. Thirty years of use and improvement of remote sensing, applied to epidemiology: From early promises to lasting frustration. *Health Place* **2007**, *13*, 400–403. [CrossRef]

11. Hay, S.I.; Packer, M.; Rogers, D. Review article The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *Int. J. Remote Sens.* **1997**, *18*, 2899–2930. [CrossRef]

12. Kalluri, S.; Gilruth, P.; Rogers, D.; Szczur, M. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: A review. *PLoS Pathog.* **2007**, *3*, e116. [CrossRef] [PubMed]

13. Hamm, N.A.S.; Soares Magalhães, R.J.; Clements, A.C.A. Earth Observation, Spatial Data Quality, and Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0004164. [CrossRef] [PubMed]

14. Zhang, Z.J.; Manjourides, J.; Cohen, T.; Hu, Y.; Jiang, Q.W. Spatial measurement errors in the field of spatial epidemiology. *Int. J. Health Geogr.* **2016**, *15*, 12. [CrossRef] [PubMed]

15. Araujo Navas, A.L.; Hamm, N.A.S.; Soares Magalhães, R.J.; Stein, A. Mapping Soil Transmitted Helminths and Schistosomiasis under Uncertainty: A Systematic Review and Critical Appraisal of Evidence. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0005208. [CrossRef] [PubMed]

16. Zhang, J.; Goodchild, M.F. *Uncertainty in Geographical Information*; Taylor & Francis Group: London, UK, 2002; ISBN 0-415-27723-X.

17. Richardson, S.; Monfort, C. Ecological Correlation Studies. In *Spatial Epidemiology: Methods and Applications*; Elliot, P.W., Wakefield, J.C., Best, N.G., Briggs, D.J., Eds.; Oxford University Press: Oxford, UK, 2000; pp. 205–220, ISBN 0192629417.

18. Wakefield, J.; Shaddick, G. Health-exposure modeling and the ecological fallacy. *Biostatistics* **2006**, *7*, 438–455. [CrossRef] [PubMed]

19. King, G. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*; Princeton University Press: Princeton, NJ, USA, 2013; ISBN 1400849209.

20. Wakefield, J.; Lyons, H. Spatial Aggregation and the Ecological Fallacy. In *Handbook of Modern Statisticcal Methods*; Hall/CRC, C., Ed.; CRC Press: Boca Raton, FL, USA, 2010; pp. 541–558, ISBN 9781420072877.

21. Gelfand, A.E.; Diggle, P.; Guttorp, P.; Fuentes, M. *Handbook of Spatial Statistics*; Taylor & Francis Group: Boca Raton, FL, USA, 2010; p. 619, ISBN 9781420072877.

22. Prentice, R.L.; Sheppard, L. Aggregate data studies of disease risk factors. *Biometrika* **1995**, *82*, 113–125. [CrossRef]

23. Richardson, S.; Stucker, I.; Hemon, D. Comparison of relative risks obtained in ecological and individual studies: Some methodological considerations. *Int. J. Epidemiol.* **1987**, *16*, 111–120. [CrossRef]

24. Wang, F.F.; Wang, J.; Gelfand, A.; Li, F. Accommodating the ecological fallacy in disease mapping in the absence of individual exposures. *Stat. Med.* **2017**, *36*, 4930–4942. [CrossRef]

25. Raso, G.; Matthys, B.; N'goran, E.; Tanner, M.; Vounatsou, P.; Utzinger, J. Spatial risk prediction and mapping of Schistosoma mansoni infections among schoolchildren living in western Côte d'Ivoire. *Parasitology* **2005**, *131*, 97–108. [CrossRef]

26. Chammartin, F.; Houngbedji, C.A.; Huerlimann, E.; Yapi, R.B.; Silue, K.D.; Soro, G.; Kouame, F.N.; N'Goran, E.K.; Utzinger, J.; Raso, G.; et al. Bayesian Risk Mapping and Model-Based Estimation of Schistosoma haematobium-Schistosoma mansoni Co-distribution in Cote d'Ivoire. *PLoS Negl. Trop. Dis.* **2014**, *8*, e3407. [CrossRef] [PubMed]

27. Woodhall, D.M.; Wiegand, R.E.; Wellman, M.; Matey, E.; Abudho, B.; Karanja, D.M.S.; Mwinzi, P.M.N.; Montgomery, S.P.; Secor, W.E. Use of Geospatial Modeling to Predict Schistosoma mansoni Prevalence in Nyanza Province, Kenya. *PLoS ONE* **2013**, *8*, e71635. [CrossRef]

28. Sturrock, H.J.W.; Pullan, R.L.; Kihara, J.H.; Mwandawiro, C.; Brooker, S.J. The Use of Bivariate Spatial Modeling of Questionnaire and Parasitology Data to Predict the Distribution of Schistosoma haematobium in Coastal Kenya. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2016. [CrossRef] [PubMed]

29. Hu, Y.; Bergquist, R.; Lynn, H.; Gao, F.; Wang, Q.; Zhang, S.; Li, R.; Sun, L.; Xia, C.; Xiong, C.; et al. Sandwich mapping of schistosomiasis risk in Anhui Province, China. *Geospat. Health* **2015**, *10*, 111–116. [CrossRef] [PubMed]

30. Souza Guimaraes, R.J.d.P.; Freitas, C.C.; Dutra, L.V.; Carvalho Scholte, R.G.; Martins-Bede, F.T.; Fonseca, F.R.; Amaral, R.S.; Drummonds, S.C.; Felgueiras, C.A.; Oliveira, G.C.; et al. A geoprocessing approach for studying and controlling schistosomiasis in the state of Minas Gerais, Brazil. *Mem. Inst. Oswaldo Cruz* **2010**, *105*, 524–531. [CrossRef]

31. Scholte, R.G.C.; Gosoniu, L.; Malone, J.B.; Chammartin, F.; Utzinger, J.; Vounatsou, P. Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models. *Acta Trop.* **2014**, *132*, 57–63. [CrossRef] [PubMed]

32. Soares Magalhães, R.J.; Salamat, M.S.; Leonardo, L.; Gray, D.J.; Carabin, H.; Halton, K.; McManus, D.P.; Williams, G.M.; Rivera, P.; Saniel, O.; et al. Mapping the Risk of Soil-Transmitted Helminthic Infections in the Philippines. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003915. [CrossRef]

33. Clements, A.C.A.; Brooker, S.; Nyandindi, U.; Fenwick, A.; Blair, L. Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *Int. J. Parasitol.* **2008**, *38*, 401–415. [CrossRef]

34. Leonardo, L.; Rivera, P.; Saniel, O.; Villacorte, E.; Lebanan, M.A.; Crisostomo, B.; Hernandez, L.; Baquilod, M.; Erce, E.; Martinez, R. A national baseline prevalence survey of schistosomiasis in the Philippines using stratified two-step systematic cluster sampling design. *J. Trop. Med.* **2012**, *2012*, 8. [CrossRef]

35. Leonardo, L.; Rivera, P.; Saniel, O.; Solon, J.A.; Chigusa, Y.; Villacorte, E.; Chua, J.C.; Moendeg, K.; Manalo, D.; Crisostomo, B.; et al. New endemic foci of schistosomiasis infections in the Philippines. *Acta Trop.* **2015**, *141*, 354–360. [CrossRef]

36. Leonardo, L.; Acosta, L.P.; Olveda, R.M.; Aligui, G.D.L. Difficulties and strategies in the control of schistosomiasis in the Philippines. *Acta Trop.* **2002**, *82*, 295–299. [CrossRef]

37. Zhou, X.N.; Bergquist, R.; Leonardo, L.; Yang, G.J.; Yang, K.; Sudomo, M.; Olveda, R. Schistosomiasis Japonica: Control and Research Needs. In *Important Helminth Infections in Southeast Asia: Diversity and Potential for Control and Elimination, Peart A*; Zhou, X.N., Bergquist, R., Olveda, R., Utzinger, J., Eds.; Elsevier Academic Press Inc.: San Diego, CA, USA, 2010; Volume 72, pp. 145–178, ISBN 978-0-12-381513-2.

38. Leonardo, L.R.; Rivera, P.; Saniel, O.; Villacorte, E.; Crisostomo, B.; Hernandez, L.; Baquilod, M.; Erce, E.; Martinez, R.; Velayudhan, R. Prevalence survey of schistosomiasis in Mindanao and the Visayas, The Philippines. *Parasitol. Int.* **2008**, *57*, 246–251. [CrossRef] [PubMed]

39. Santos, F.L.N.; Cerqueira, E.J.L.; Soares, N.M. Comparison of the thick smear and Kato-Katz techniques for diagnosis of intestinal helminth infections. *Rev. Soc. Bras. Med. Trop.* **2005**, *38*, 196–198. [CrossRef] [PubMed]

40. ESRI. *ArcGIS Desktop, 10*; Environmental Systems Research Institute: Redlands, CA, USA, 2011.

41. Zhou, Y.B.; Liang, S.; Jiang, Q.W. Factors impacting on progress towards elimination of transmission of schistosomiasis japonica in China. *Parasit. Vectors* **2012**, *5*, 7. [CrossRef] [PubMed]

42. Walz, Y.; Wegmann, M.; Dech, S.; Raso, G.; Utzinger, J. Risk profiling of schistosomiasis using remote sensing: Approaches, challenges and outlook. *Parasit. Vectors* **2015**, *8*, 16. [CrossRef] [PubMed]

43. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [CrossRef]

44. Prah, S.; James, C. The influence of physical factors on the survival and infectivity of miracidia of Schistosoma mansoni and S. haematobium I. Effect of temperature and ultra-violet light. *J. Helminthol.* **1977**, *51*, 73–85. [CrossRef]

Int. J. Environ. Res. Public Health **2019**, *16*, 176

16 of 17

45. Woolhouse, M.; Chandiwana, S. Population dynamics model for Bulinus globosus, intermediate host for Schistosoma haematobium, in river habitats. *Acta Trop.* **1990**, *47*, 151–160. [CrossRef]

46. Pietrock, M.; Marcogliese, D.J. Free-living endohelminth stages: At the mercy of environmental conditions. *Trends Parasitol.* **2003**, *19*, 293–299. [CrossRef]

47. Geological Survey, U.S. Global Data Explorer. Available online: https://gdex.cr.usgs.gov/gdex/ (accessed on 15 August 2017).

48. Pesigan, T.P.; Hairston, N.G.; Jauregui, J.J.; Garcia, E.G.; Santos, A.T.; Santos, B.C.; Besa, A.A. Studies on Schistosoma japonicum infection in the Philippines 2. The molluscan host. *Bull. World Health Organ.* **1958**, *18*, 481–578.

49. Stensgaard, A.S.; Jorgensen, A.; Kabatereine, N.B.; Rahbek, C.; Kristensen, T.K. Modeling freshwater snail habitat suitability and areas of potential snail-borne disease transmission in Uganda. *Geospat. Health* **2006**, *1*, 93–104. [CrossRef] [PubMed]

50. Stensgaard, A.S.; Utzinger, J.; Vounatsou, P.; Hurlimann, E.; Schur, N.; Saarnak, C.F.L.; Simoonga, C.; Mubita, P.; Kabatereine, N.B.; Tchuente, L.A.T.; et al. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: Does climate matter? *Acta Trop.* **2013**, *128*, 378–390. [CrossRef] [PubMed]

51. National Mapping and Resource Information Authority. Available online: http://www.namria.gov.ph/ (accessed on 15 February 2018).

52. Gelman, A. Prior distributions for variance parameters in hierarchical models (Comment on an Article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–533. [CrossRef]

53. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1995; ISBN 1135439419.

54. Diggle, P.J.; Tawn, J.; Moyeed, R. Model-based geostatistics. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2002**, *47*, 299–350. [CrossRef]

55. Thomas, A.; Best, N.; Lunn, D.; Arnold, R.; Spiegelhalter, D. *GeoBugs User Manual*; MRC Biostatistics Unit: Cambridge, UK, 2004.

56. Lunn, D.; Jackson, C.; Best, N.; Thomas, A.; Spiegelhalter, D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*; CRC Press: Boca Raton, FL, USA, 2012; ISBN 1584888490.

57. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434–455. [CrossRef]

58. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **1992**, *7*, 457–472. [CrossRef]

59. Brooker, S.; Hay, S.I.; Bundy, D.A.P. Tools from ecology: Useful for evaluating infection risk models? *Trends Parasitol.* **2002**, *18*, 70–74. [CrossRef]

60. Hijmans, R.; Rojas, E.; Cruz, M.; O'Brien, R.; Barrantes, I. DIVA-GIS free, simple and effective. Available online: http://www.diva-gis.org/Data (accessed on 8 April 2018).

61. Project, O.S.M. *Planet OSM*; Open Street Map Project: Cambridge, UK, 2017.

62. OCHA Humanitarian Data Exchange v1.25.3. Available online: https://data.humdata.org/search?groups=phl&q=&ext_page_size=25 (accessed on 10 April 2018).

63. Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. *OpenBUGS User Manual*; Version 3.0.2.; MRC Biostatistics Unit: Cambridge, UK, 2007.

64. Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. *WinBUGS User Manual*; MRC Biostatistics Unit: Cambridge, UK, 2003.

65. Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. OpenBUGS license. Available online: http://www.openbugs.net/w/Downloads (accessed on 25 June 2018).

66. Sturtz, S.; Ligges, U.; Gelman, A. *R2OpenBUGS: A Package for Running OpenBUGS from R.*; Journal of Statistical Software: Innsbruck, Austria, 2010.

67. Team, R.D.C. *R: A Language and Environment for Statistical Computing*; The R Foundation for Statistical Computing: Vienna, Austria, 2013.

68. Brooker, S.; Hay, S.; Issae, W.; Hall, A.; Kihamia, C.; Lawambo, N.; Wint, W.; Rogers, D.; Bundy, D. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Trop. Med. Int. Health* **2001**, *6*, 998–1007. [CrossRef]

69. Kristensen, T.; Malone, J.; McCarroll, J. Use of satellite remote sensing and geographic information systems to model the distribution and abundance of snail intermediate hosts in Africa: A preliminary model for Biomphalaria pfeifferi in Ethiopia. *Acta Trop.* **2001**, *79*, 73–78. [CrossRef]

70. Brooker, S.; Hay, S.I.; Tchuente, L.-A.T.; Ratard, R. Using NOAA-AVHRR data to model human helminth distributions in planning disease control in Cameroon, West Africa. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 175–179.

71. Malone, J.B.; Yilma, J.M.; McCarroll, J.C.; Erko, B.; Mukaratirwa, S.; Zhou, X.Y. Satellite climatology and the environmental risk of Schistosoma mansoni in Ethiopia and east Africa. *Acta Trop.* **2001**, *79*, 59–72. [CrossRef]

72. Rice Science for a Better World. Available online: http://irri.org/our-work/research/policy-and-markets/mapping-rice-in-the-philippines-where (accessed on 18 October 2018).

73. Stensgaard, A.; Jorgensen, A.; Kabatareine, N.; Malone, J.; Kristensen, T. Modeling the distribution of Schistosoma mansoni and host snails in Uganda using satellite sensor data and Geographical Information Systems. *Parassitologia* **2005**, *47*, 115–125. [PubMed]

74. Burns, C.J.; Wright, M.; Pierson, J.B.; Bateson, T.F.; Burstyn, I.; Goldstein, D.A.; Klaunig, J.E.; Luben, T.J.; Mihlan, G.; Ritter, L.; et al. Evaluating Uncertainty to Strengthen Epidemiologic Data for Use in Human Health Risk Assessments. *Environ. Health Perspect.* **2014**, *122*, 1160–1165. [CrossRef]

75. Schur, N.; Hurlimann, E.; Garba, A.; Traore, M.S.; Ndir, O.; Ratard, R.C.; Tchuente, L.A.T.; Kristensen, T.K.; Utzinger, J.; Vounatsou, P. Geostatistical Model-Based Estimates of Schistosomiasis Prevalence among Individuals Aged <= 20 Years in West Africa. *PLoS Negl. Trop. Dis.* **2011**, *5*, e1194. [CrossRef] [PubMed]

76. Schur, N.; Hurlimann, E.; Stensgaard, A.S.; Chimfwembe, K.; Mushinge, G.; Simoonga, C.; Kabatereine, N.B.; Kristensen, T.K.; Utzinger, J.; Vounatsou, P. Spatially explicit Schistosoma infection risk in eastern Africa using Bayesian geostatistical modelling. *Acta Trop.* **2013**, *128*, 365–377. [CrossRef] [PubMed]

77. Simoonga, C.; Utzinger, J.; Brooker, S.; Vounatsou, P.; Appleton, C.C.; Stensgaard, A.S.; Olsen, A.; Kristensen, T.K. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology* **2009**, *136*, 1683–1693. [CrossRef] [PubMed]

78. White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **1980**, 817–838. [CrossRef]

79. Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods*; Sage Publications, Inc.: Thousand Oaks, CA, USA, 1997; ISBN 080394540X.

80. Mankiw, N.G. A Quick Refresher Course in Macroeconomics. *J. Econ. Lit.* **1990**, *28*, 1645–1660.