Parasites & Vectors

RESEARCH

Open Access

CrossMark

Modelling local areas of exposure to *Schistosoma japonicum* in a limited survey data environment

Andrea L. Araujo Navas[1*], Ricardo J. Soares Magalhães[2,3], Frank Osei[1], Raffy Jay C. Fornillos[4], Lydia R. Leonardo[5] and Alfred Stein[1]

Abstract

Background: Spatial modelling studies of schistosomiasis (SCH) are now commonplace. Covariate values are commonly extracted at survey locations, where infection does not always take place, resulting in an unknown positional exposure mismatch. The present research aims to: (i) describe the nature of the positional exposure mismatch in modelling SCH helminth infections; (ii) delineate exposure areas to correct for such positional mismatch; and (iii) validate exposure areas using human positive cases.

Methods: To delineate exposure areas to *Schistosoma japonicum*, a spatial Bayesian network (sBN) was constructed. It uses data on exposure risk factors such as: potential sites for snails' accessibility, geographical distribution of snail infection rate, and cost of the community to access nearby water bodies. Prior and conditional probabilities were obtained from the literature and inserted as weights based on their relative contribution to exposure; these probabilities were then used to calculate joint probabilities of exposure within the sBN.

Results: High values of probability of *S. japonicum* exposure correspond to polygons where snails could potentially be present, for instance in wet soils and areas with low slopes, but also where people can easily access water bodies. Low correlation ($R^2 = 0.3$) was found between the percentage of human cases and the delineated probabilities of exposure when validation buffers are generated over the human cases.

Conclusions: The utility of a probabilistic method for the identification of exposure areas for *S. japonicum*, with wider application for other water-borne infections, was demonstrated. From a public health perspective, the schistosomiasis exposure sBN developed in this study could be used to guide local schistosomiasis control teams to specific potential areas of exposure, and improve efficiency of mass drug administration campaigns in places where people are likely to be exposed to the infection.

Keywords: Schistosomiasis, Spatial modeling, Bayesian network, Exposure uncertainty, Risk factors

Background

Schistosomiasis (SCH) is a water-borne neglected tropical disease of global public health significance [1, 2]. It affects more than 252 million people worldwide [3], especially human populations living in places where clean water and sanitation are limited [4]. Schistosomiasis is known to lead to anaemia, stunted growth and other organ pathologies in school-aged children [5, 6]. Three schistosome species cause the infection: *Schistosoma mansoni, S. japonicum* and *S. haematobium*. *Schistosoma japonicum* is presently endemic in China, Indonesia and the Philippines, and is hard to control due to its zoonotic life-cycle [7]. The life-cycle of *S. japonicum* includes infection of an amphibious snail belonging to several subspecies of *Oncomelania hupensis* as the intermediate host, and humans and other mammalians as definitive hosts [8, 9].

Traditionally, schistosomiasis risk mapping has enabled the identification of at risk populations for targeting mass drug administration campaigns, thus increasing

* Correspondence: a.l.araujonavas@utwente.nl
[1]Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands
Full list of author information is available at the end of the article

Araujo Navas et al. Parasites & Vectors (2018) 11:465

Page 2 of 15

the efficiency of schistosomiasis disease control [10]. Schistosomiasis mapping has been supported by the use of spatial information techniques, such as geographical information systems (GIS), remote sensing and global positioning systems (GPS). Spatial information techniques allow the manipulation of spatially referenced infection data and data on the physical and biological environmental variables [11–14]. Modelling those data in combination allows studying the distribution of communities most at risk schistosomiasis and the role of the geographical variation of environmental exposure factors on schistosomiasis risk [15].

There are a number of errors inherent to spatial information used in geographical epidemiological studies [4]. Most of these errors involve positional measurement errors, where observation and prediction locations are affected by various factors such as GPS inaccuracies, the presence of multiple addresses, geocoding errors, outcome or covariate aggregations, and misalignment between covariates of exposure and disease outcome estimates [15]. The last one is of our current interest and may occur when covariates of exposure are extracted from locations where exposure has not occurred.

Statistical modelling of the spatial distribution of schistosome infections estimates empirical relationships between morbidity indicators (e.g. prevalence or intensity of infection) and risk factors. Risk factors for schistosome exposure include various environmental and socio-economic covariates that help to interpolate the level of infection at unsampled locations [14, 16–18]. Covariates and morbidity indicators are commonly extracted from survey locations such as health centres, hospitals and schools. In most cases, exposure to infection did not occur at survey data locations but at locations where environmental and geographical conditions, together with the level of accessibility to contaminated sites, are optimally exposed. Such exposure locations are usually unknown, resulting in positional mismatch of the surveyed disease values, and the covariates in the model.

To date, methods to account for this type of positional misalignment are scarce. Several studies have used remote sensing data to determine biophysical features of habitats in relation to snail prevalence [19–24], acknowledging that S. japonicum transmission is closely related to the distribution of its intermediate host in the environment [9]. Only one study [2] has used these habitats to correct for the positional mismatch when modelling disease infection risk in human populations. Walz et al. [2] used high-resolution remote sensing data, environmental field measurements, and ecological data, to model environmental suitability for schistosomiasis-related parasites and snail species. They represented environmental suitability as potential transmission areas that could guide public health interventions to places where people could potentially be infected. Although

potential transmission areas were delineated, interactions between humans, hosts, and suitable environments were not taken into account.

These studies suggest that ignoring positional mismatch and its impact on spatial prediction remains largely unquantified in schistosomiasis modelling. Furthermore, the extraction of covariate values in the presence of positional mismatch is a significant source of uncertainty that may influence the efficacy of schistosomiasis control strategies [4]. Therefore, methods to correct for this positional mismatch need to be further investigated [1, 4].

The objective of this study is to develop a schistosomiasis exposure sBN model that maps potential areas of exposure to S. japonicum, taking into account human interactions with main sources of infection (i.e. water bodies). To accomplish this objective, we aimed to (i) describe the positional mismatch problem in modelling S. japonicum infection; (ii) delineate exposure areas that take into consideration the accessibility cost of people to main sources of infection, and that could be used to correct for this positional mismatch; and to (iii) validate the delineated exposure areas.

Methods

Data on human and snail S. japonicum infection

In the Philippines S. japonicum is endemic in 28 of its 81 provinces [25], with approximately 1.8 million estimated infected people [26]. The disease affects children, adolescents and individuals with high-risk occupations, such as farmers and fishermen [26, 27]. In the Philippines, the smallest administrative division is the barangay, numbering about 22–50 in a municipality.

We used data on human schistosomiasis and snail prevalence of infection, collected in six barangays from Alangalang municipality in Leyte Province in 2015 and 2016. Data were collected by researchers from the College of Public Health and College of Science from the University of the Philippines. Surveyors selected Alangalang municipality because it has the highest prevalence of schistosomiasis (7.5%) from all the 43 municipalities of Leyte Province; within this municipality, they visited the barangays with the highest prevalence of infection from the 54 barangays in Alangalang municipality.

Human positive cases (12 records) were georeferenced at household locations and snails surveys (8 records) were taken from water bodies in close proximity to surveyed households. The recording of all the human case locations (also including negative cases) was not possible due to a lack of manpower and material resources, such as the availability of only one GPS device in the field.

Diagnosis of schistosomiasis in humans was performed using stool examination. Single stool sample was requested per participant with informed consent, coded

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 3 of 15

and prepared following the Kato-Katz method. Each slide prepared was read in the field using a microscope and the presence of *S. japonicum* eggs indicated active infection.

Infection among *O. h. quadrasi* snails was determined by manually crushing the snails in aliquots on a glass slide. Each snail was placed in an aliquot droplet of distilled water, usually three aliquots per glass slide. Snails were gently crushed in between slides and were examined under a conventional stereomicroscope (40×) using forceps for separating snail tissues to detect the presence of sporocysts or furcocercous cercariae characteristic of *S. japonicum*.

Study area
For the purpose of this study, it was decided to work at a local spatial scale in the Province of Leyte, due to the localized nature of the surveys and the high endemicity of the disease [28]. For the analysis, we identified a small area surrounding surveyed points (Fig. 1). This was done in order to select only surveyed barangays and to include information of all risk factors, avoiding areas without survey information (Fig. 1).

Environmental and geographical data
Exposure risk factors of SCH transmission are associated with the environment (i.e. moisture, temperature, rainfall and water characteristics), the topography (i.e. elevation, slope) of the area [2, 10, 20, 21, 29] and snail infection status [23, 24, 30, 31]. In the endemic provinces of the Philippines, exposure to snails is mostly driven by the local topography, land use and the physical and chemical components of the water and soil [32]. We included elevation, slope, land use, nearest distance to water bodies and snail infection rates as exposure risk factors. Elevation was obtained as a raster file from Aster GDEM version 2 from USGS [33]. Vector layers for land use, river and road network were obtained from the OpenStreetMap (OSM) project [34]. OSM land use and land cover products use information from GlobeLand30 (GL30), which is a new generation of 30 m land cover maps [35–37]. The OSM road and river networks are incomplete and contain errors in their connectivity. To account for this, we edited roads and rivers, and digitalized footpaths using Google Earth images. The vector layer for snail infection rate was obtained from the recorded surveys (Table 1). Slope was derived from elevation by using the Terrain Analysis tool from Quantum GIS version 2.6 [38].

Distance to water bodies was calculated using the closest facility network analysis tool from ArcGIS version 10 [39]. Firstly, we corrected for topology errors such as duplicate lines, presence of dangles and multipart geometries in the river and road network. Secondly, communities were loaded as incidents (261 points), and contact

river points as facilities (42 points). Thirdly, we used the closest facility tool to find the nearest river from an urban area following a road. Finally, we interpolated the distance to the nearest water source using ordinary kriging from the *gstat* package in R [40] and saved the map as a raster file.

Snail infection rate map
We constructed a trend surface that represents snail infection rate for the whole study area, thus using data of all the points to predict at unknown locations (i.e. global interpolation). It fits a mathematically defined surface through the data points (i.e. deterministic interpolation) to discover smoother (i.e. inexact interpolation) regional and local trends. It is similar to a three dimensional regression surface obtained with linear regression, where coordinates $s_i = (x_i, y_i)$ are used as predictors. The interpolated value $z(S_i)$ for a first and second order polynomial is represented in equations 1 and 2, respectively. $z(S_i)$ represents infection rate values (number of positive cases/number of sampled snails) at location i.

$$z^*(s_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i \tag{1}$$

$$z^*(s_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i \tag{2}$$

Figure 2a, b shows the resulting surfaces for the first and second order polynomials, respectively. Figure 2a shows low risk probability values (Table 1), from -0.003 to 0.008. These values do not match the original surveyed values. Figure 2b shows low and medium risk probability values from -0.01 to 0.035. These values show a better fit to the original surveyed values showed in red.

To remove the occurring negative values, we fitted a multiple linear regression by applying a generalized linear regression model using equation 2. In this case $z(S_i)$ was the infection status for each location i, 1 indicates an infected case and 0 a non-infected case. The resulting prediction from Fig. 3 shows only positive predicted values but very large standard errors (28.7 to 3e+13). Besides, none of the predictions approximate the original surveyed values. Finally, the second order trend surface (Fig. 2b) map was used for the analysis since it better fitted the original surveyed values.

Spatial Bayesian network of *Schistosoma japonicum* exposure
We have conceptually designed a model that represents the positional mismatch between survey locations and
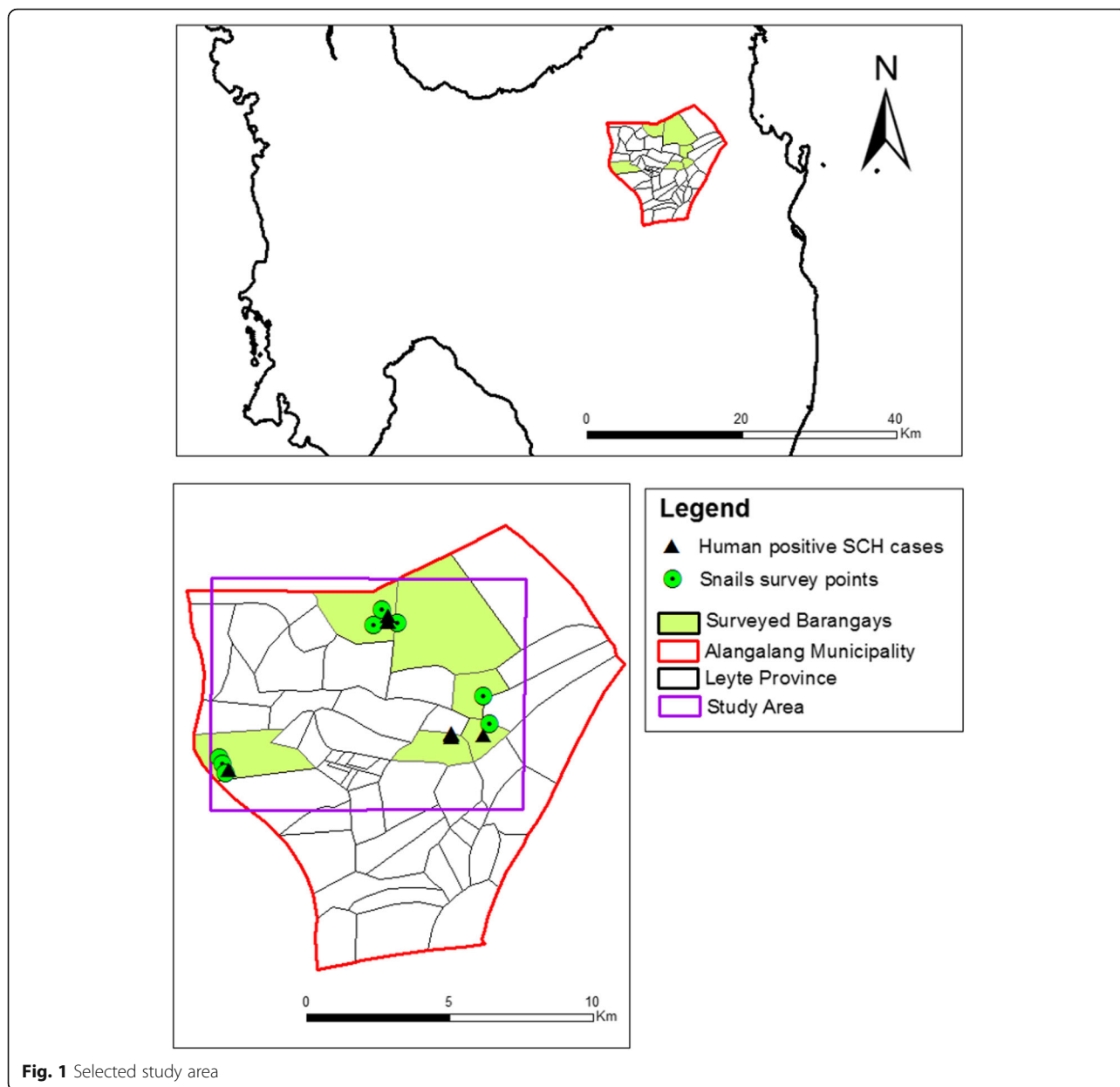
Fig. 1 Selected study area

exposure sites (Fig. 4). Locations s_1 and s_2 represent the schools, households, or other survey locations from which morbidity indicators are extracted, while ex_{mn} represents the various exposure points where infections could have taken place, m is the corresponding number of exposure points and n is the corresponding survey locations related to the exposure.

Exposure areas were delineated by using spatial Bayesian networks (sBN) [41]. A Bayesian network (BN) is a probabilistic graphical model that captures the various conditional dependencies of a set of random variables (discrete or continuous) [42, 43], into a joint probability distribution by means of a directed acyclic graph (DAG) [44, 45]. A BN for a set of random variables X is defined

by the pair (D, P). Here, D is the DAG and P is the set of probability distributions for all variables in the network. Each variable x with parents $pa(x)$ has a conditional probability $p(x| \, pa(x))$. For a BN with a set of discrete (I) variables, the joint probability distribution factorizes into equation 3 [42]. This is the joint probability distribution as the product of all conditional probabilities specified in a BN:

$$p(X) = \prod_{i=1}^{I} p(x_i | pa(x_i)) \tag{3}$$

The schistosomiasis exposure sBN defines exposure areas in a probabilistic way, by allowing the combination
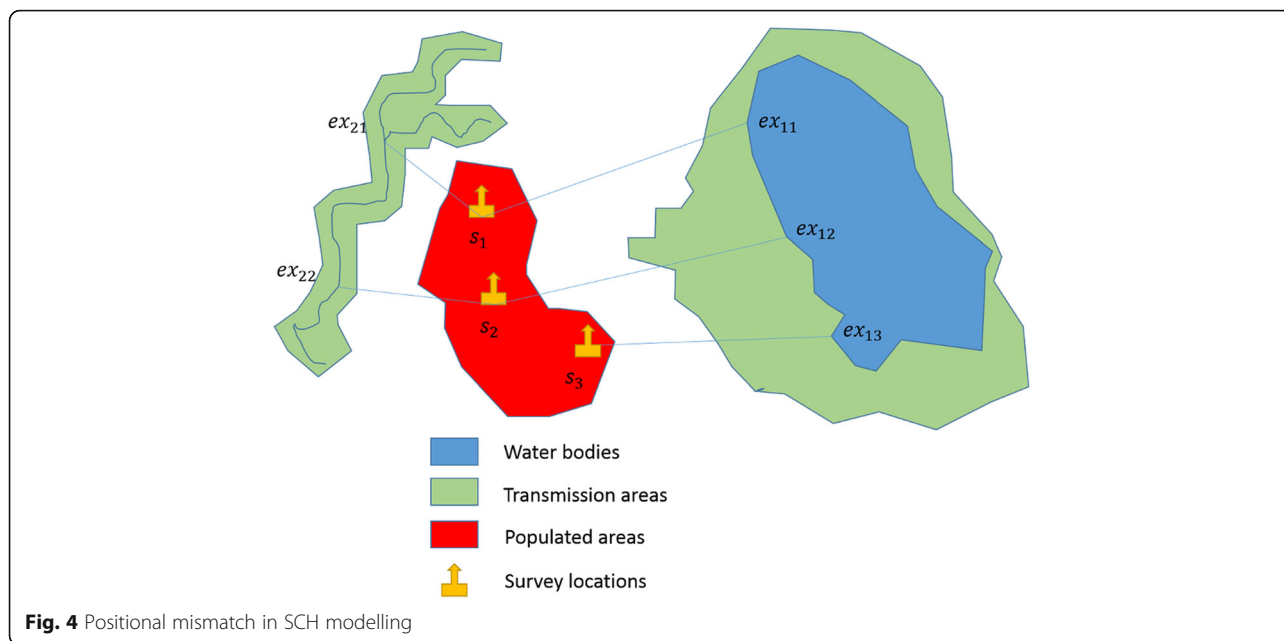
Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 5 of 15

Table 1 Categorization of exposure risk factors

Risk factor (weight)	Spatial resolution	Temporal resolution	Data type	Coordinate system	Data source	Hypothetical link	Classification	π weights	Based upon
Elevation (0.03)	~ 30 m at equator	na	Raster	EPSG:4326	Aster GDEM V2 from USGS	While elevation decreases, the risk of infection increases	High risk: < 900 m	0.70	[32, 51, 56]
							Medium risk: 900–2300 m	0.25	
							Low risk: > 2300 m	0.05	
Land use (0.26)	~ 30 m	2-3-2017	Vector	EPSG:4326	OpenStreetMap project	Wet surfaces are more suitable to a higher risk of infection	Very high risk: wet soils	0.42	[32, 57]
							High risk: water bodies	0.29	
							High and medium risk: Agriculture land and grass	0.16	
							Medium and low risk: forest and natural areas	0.08	
							Low risk: barren land	0.02	
							Very low risk built land	0.03	
Slope (0.13)	~ 30 m at equator	na	Raster	EPSG:4326	Derived from elevation	At more flat surfaces the risk of infection increases	High risk: < 11 degrees	0.70	[49, 51]
							Medium risk: 11–30 degrees	0.23	
							Low risk: > 30 degrees	0.07	
Distance to water bodies (0.50)	30 m	2-3-2017	Raster	EPSG:32651	Derived from roads, urban areas, river network and water bodies from the OpenStreetMap project	While distance to water bodies decreases, the risk of infection increases	High risk: < 1000 m	0.74	[51, 52, 58]
							Medium risk: 1000–5000 m	0.21	
							Low risk: > 5000 m	0.05	
Snail infection rate (0.06)	na	2015–2016	Vector	EPSG:4326	Derived from recorded surveys	While snail infection rate increases, the risk of infection increases	High risk: > 3.6%	0.65	[23, 24, 30, 31]
							Medium risk: 0.5–3.6%	0.28	
							Low risk: < 0.5%	0.07	

Abbreviation: na, not applicable

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 6 of 15

Fig. 2 First order (**a**) and second order (**b**) polynomial trend surface. Red crosses represent the original surveyed snail infection locations

of various probability distributions from a set of random spatial variables [44]. We have constructed a DAG for exposure areas (Fig. 5), where each random variable is represented as a node. Nodes are connected by directed links or edges that express probabilistic relationships between the variables [43]. Three types of random variables can be found including (i) an observable discrete variable [land use (*LU*)]; (ii) observable continuous variables [elevation (*E*), slope (*SLP*), distance to water bodies (*DWB*) and snail infection rates (*SI*)]; and (iii) latent

discrete variables [potential accessible sites for snails (*PAS*), community cost (*CC*) and exposure (*EX*)]. The direction given in the link between variables, for instance from *LU* to *PAS*, encodes a direct causal dependence of *PAS* on *LU*; the node *LU* is known then as the parent of *PAS* [45].

All continuous variables (E, SLP, DWB and SI) were discretized into different categories, given that high or low levels of exposure could occur at various ranges of risk factor values. We established hypothetical relationships between the risk factors and the disease, and categorized the risk factors based on literature (Table 1).

Exposure is a discrete child node, which has three discrete parent nodes: *PAS*, *CC* and *SI*; its conditional probability is expressed as $p(EX| PAS, CC, SI)$. *PAS* and *CC* are at the same time child nodes conditional on discrete parents. Their conditional probabilities are derived by $p(PAS| LU, E, SLP)$ and $p(CC| DWB)$, respectively. The joint probability distribution for our Bayesian network is given as:

$$p(X) = p(EX|PAS, CC, SI) \cdot p(PAS|LU, E, SLP) \cdot$$
$$p(LU) \cdot p(E) \cdot p(SLP) \cdot p(CC|DWB) \cdot p(DWB) \cdot p(SI)$$
$$(4)$$

Equation 4 encodes assumptions of this research about direct dependencies between variables and indicates which node probability tables (NPT) need to be defined [45].

Fig. 3 Predicted probability of snail infection values using generalized linear regression model. Colour scale represent probability values from 0 to 1. Snail survey locations are represented by white crosses

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 7 of 15

Fig. 4 Positional mismatch in SCH modelling

Construction of node probability tables

After defining the structure of our sBN, a main challenge is to construct the node probability tables (NPT). NPT are probability tables associated to each child node v given every possible state of the set of parents of v. NPT are intended to capture the strength of the relationship between the node and its parents [45]. The practicality of doing this depends on the number of states of the parent and child nodes. In our sBN eight NPTs were constructed, five NPTs as prior marginal probabilities (π) were inserted for the set of parent nodes (LU, E, SLP, DWB and SI) and three NPTs as conditional probabilities linking parent and child nodes (PAS, CC and EX).

We inserted prior marginal probabilities for the set of discrete parent nodes as weights. Weights were calculated using the eigen vector derived from a pairwise comparison matrix using Saaty's comparison table [46]. Saaty [46] uses a scale of numbers (i.e. scale of judgement) to indicate how many times a factor is more dominant than another with respect to a criterion used for their comparison. In this case, the criterion is the risk of infection assigned to each parent node category given by literature (Table 1). Consistency indexes and ratios were calculated in order to measure the consistency of the judgements. Consistency ratios lower than 10%, indicate that our judgements are acceptable, while consistency ratios higher than 10% indicate untrustworthy judgements or random decisions. Saaty's pairwise matrices as
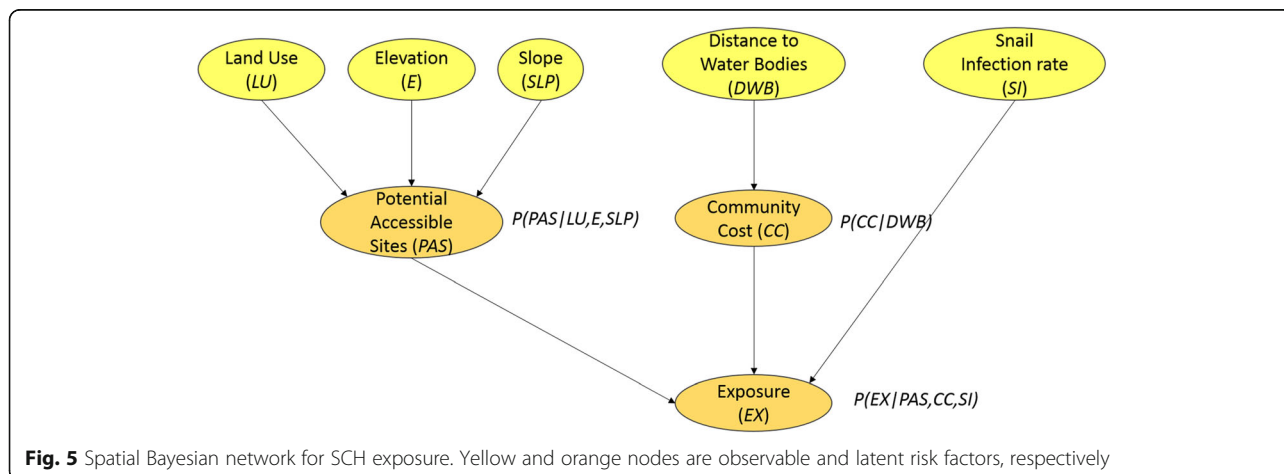
Fig. 5 Spatial Bayesian network for SCH exposure. Yellow and orange nodes are observable and latent risk factors, respectively

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 8 of 15

well as consistency indexes and ratios are included as Additional file 1: Tables S1-S7. Prior marginal probabilities for the parent nodes are shown in Table 1.

Latent variables *PAS, CC* and *EX* were divided into three probability categories: high, medium and low risk. Conditional probabilities for these child nodes are associated with the edges that link them to the parent nodes, and were also assigned using a pairwise comparison matrix. The criterion used to assign the scale of judgement is the strength of the hypothetical link between the risk factors and exposure. The strength of the hypothetical link was evaluated based upon three studies that evaluated the risk factors associated with schistosomiasis infection [47–49].

Hu et al. [47] ranked the potential importance of the schistosomiasis risk factors by means of a power detector. According to this detector, distance to water bodies is the most significant factor for disease risk, and elevation the least significant. Zhang et al. [48] used environmental, topographical and human behavioural factors to locate schistosomiasis active transmission sites. Their predictor capacity was compared by means of deviance analysis, used to determine the important variables to be included in a generalized additive model. As in the previous study, distance to water bodies was the most significant factor because of the smallest deviance, and elevation the least significant. Finally, Ajakaye et al. [49] evaluated physical and environmental risk factors to identify areas with suitable conditions for schistosomiasis transmission. They used Saaty's comparison matrix to assign weights to each risk factor. Distance to water bodies and land use were the most significant factors, followed by elevation and slope as the least significant.

Weights obtained for each risk factor are shown in Table 1 and the conditional probabilities linking parent and child nodes are shown in Additional file 2: Tables S8-S10.

Deriving joint probabilities
To compute the probabilities for each category of the child nodes, *PAS, CC* and *EX*, conditional and marginal probabilities were used by applying equations 5, 6 and 7, respectively. Joint probability values of exposure were calculated for each polygon of analysis. In order to update the prior marginal probabilities, evidence is inserted for each spatial polygon into the observable variables (*SI, LU, E, SLP, DWB*). Bold facing indicates the insertion of evidence. Variables notation can be found in Additional file 3: Table S11.

$$p(PAS, \boldsymbol{LU}, \boldsymbol{E}, \boldsymbol{SLP}) = \sum\nolimits_{LU,E,SLP} p(\boldsymbol{LU}) * p(\boldsymbol{E}) * p(\boldsymbol{SLP}) * p(PAS|\boldsymbol{LU},\boldsymbol{E},\boldsymbol{SLP})$$

(5)

$$p(CC, \boldsymbol{DWB}) = \sum\nolimits_{DWB} p(\boldsymbol{DWB}) * p(CC|\boldsymbol{DWB})$$

(6)

$$p(EX, PAS, CC, \boldsymbol{SI}) = \left(\sum\nolimits_{PAS,CC,SI} p(EX|PAS, CC, \boldsymbol{SI}) \right.$$
$$* p(\boldsymbol{SI}) \left(\sum\nolimits_{LU,E,SLP} p(\boldsymbol{LU}) * p(\boldsymbol{E}) * p(\boldsymbol{SLP}) \right.$$
$$\left. * p(PAS|\boldsymbol{LU},\boldsymbol{E},\boldsymbol{SLP}) \right) \left(\sum\nolimits_{DWB} p(\boldsymbol{DWB}) * p(CC|\boldsymbol{DWB}) \right) \right)$$

(7)

For the implementation, polygons of analysis were constructed based on the overlaying of each risk factor (i.e. parent node). To overlay all risk factors, they were first transformed into vectors and then corrected for topology errors. Topology errors included duplicated polygons, multipart geometries and overlapping polygons.

Sensitivity analysis was used to see the relative influence of the risk factors on *PAS* and *CC*, and the relative influence of *PAS, CC* and *SI* on exposure. We used the sensitivity function, calculated as the degree of entropy reduction. Degree of entropy reduction *I* is the degree of change or expected difference in information bits *H* between a query variable *Q* (exposure) with *q* states and findings variable *F* (risk factors) with *f* states [50] (equation 8). A degree of entropy reduction of 0 means a query variable is independent of the varying variable.

$$\{I = H(Q) - H(F) = \sum_q \sum_f \frac{P(q,f) \log_2[P(q,f)]}{P(q)P(f)}$$

(8)

Software
To work within the spatial domain we used the software Netica$^{\text{TM}}$ 6.03 [41], which works with Bayesian networks, decision nets and influence diagrams. Evidence is inserted as cases for each polygon of analysis, and prior and conditional probabilities are inserted as tables.

Validation
Validation was first performed by counting all surveyed positive SCH human cases falling inside the various categories of exposure in the map. However, this introduces a positional mismatch as the surveyed positive cases were not necessarily acquired at those specific exposure points.

As a second approach for validation, we defined potential validation areas by constructing buffers around each of the positive cases. We extracted the distance to the nearest water body for each surveyed point using the distance map previously generated. Extracted distance values were used as distance buffers generated around

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 9 of 15

positive cases. Buffers completely containing other buffers were grouped. We counted the number of positive cases falling inside each group and calculated the mean probability of exposure within the grouped buffers.
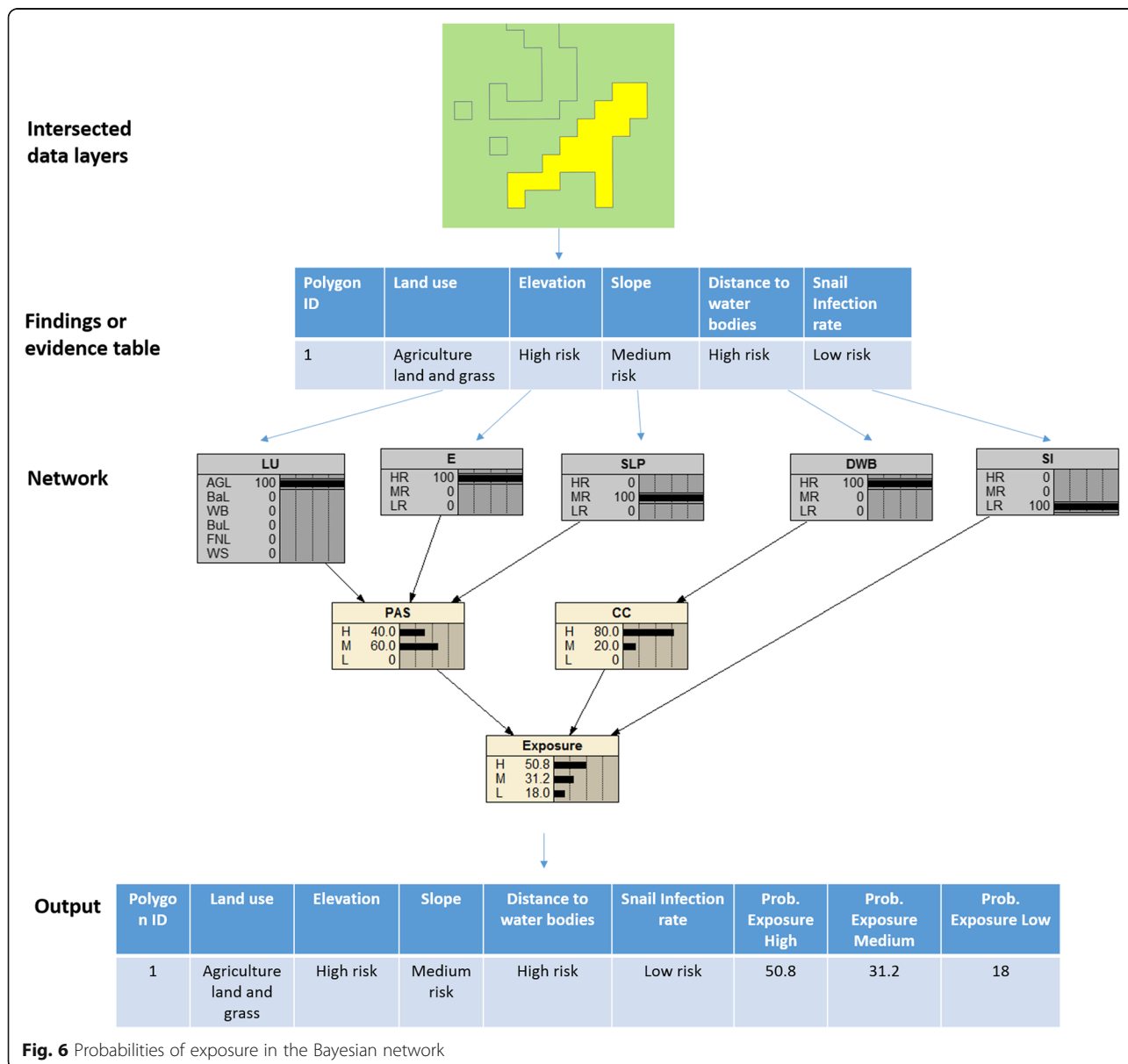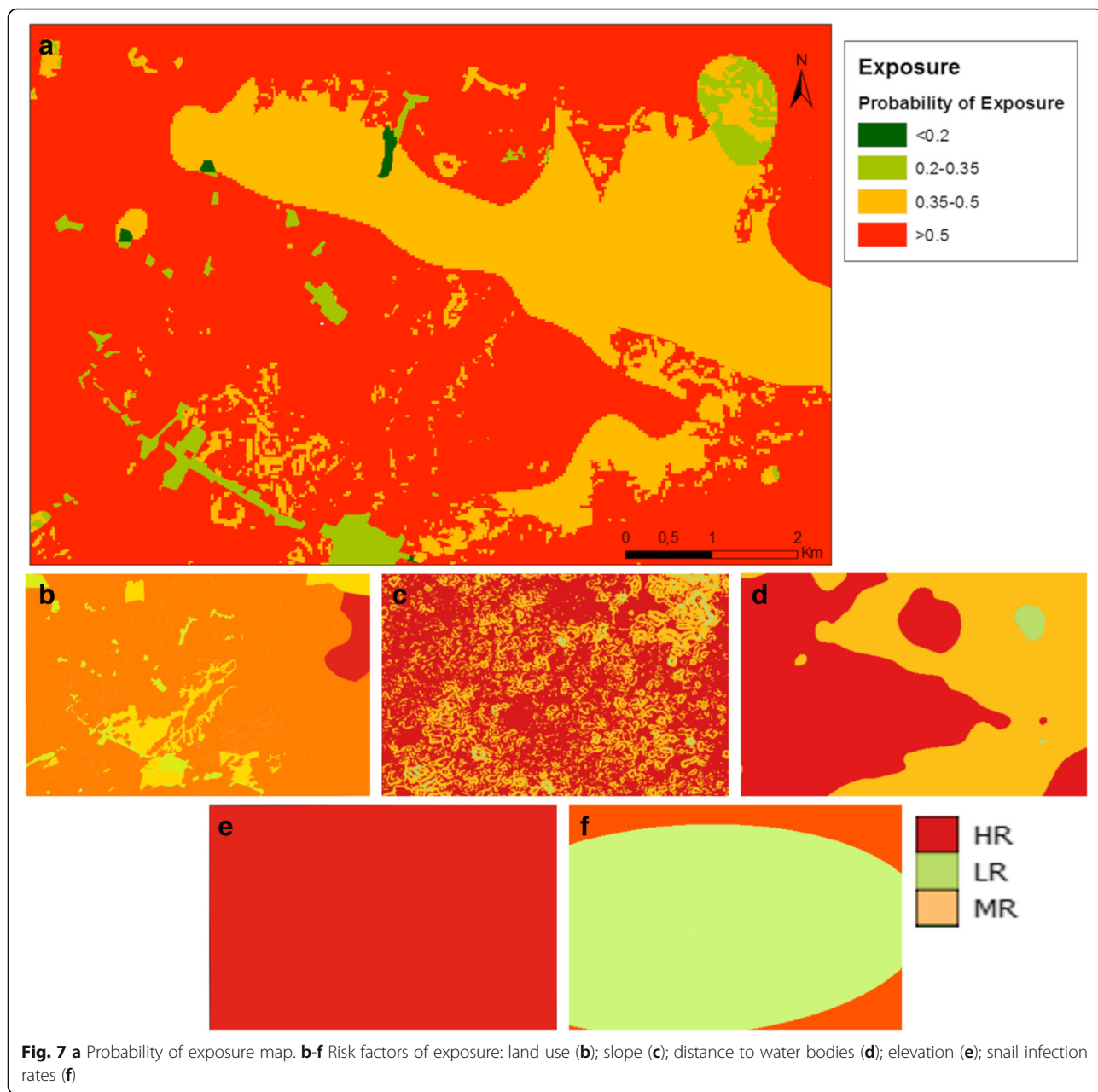
Results

Exposure network

High (> 50%), medium (35–50%), low (20–35%) and very low (< 20%) probabilities of exposure were derived from the proposed exposure network. This is exemplified in Fig. 6 for only one polygon. For this particular polygon, the probability is predominantly high (50.8%) for a high-risk elevation (< 900 m), *DWB* (< 1 km), and *LU*

(agriculture land and grass), a medium risk slope (11–30°), and a low risk *SI* (< 0.5%).

Very low probability values of exposure (< 20%) were found in built-up areas, medium risk *DWB* (1–5 km), slopes < 30° and low and medium (0.5–3.6%) risk of snail infection, but also in agriculture and grass land with *DWB* > 5 km and slopes > 30° (Fig. 7). Low probabilities of exposure (20–35%) were found in built-up areas with slopes < 30°, low risk of snail infection, and within *DWB* < 1 km, but also in agriculture and grass land in *DWB* > 5 km. Medium probability values (35–50%) were found in agriculture and grass land and forest areas, in slopes > 11°, low risk of snail infection, and *DWB* < 1 km, but also in slopes < 30°, medium risk of snail infection and *DWB* from 1 to 5 km. High

Fig. 6 Probabilities of exposure in the Bayesian network

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 10 of 15

Fig. 7 a Probability of exposure map. **b-f** Risk factors of exposure: land use (**b**); slope (**c**); distance to water bodies (**d**); elevation (**e**); snail infection rates (**f**)

probability of exposure values (> 50%) were found in wet soils with slopes < 30°, with *DWB* from 1 to 5 km and medium risk of snail infection, but also in agriculture and grass land with *DWB* < 1 km and low risk of snail infection.

Based on the degree of entropy reduction, our sensitivity results show that the risk factor with the highest degree of change is *PAS* followed by *SI* and *CC*. Within PAS, land use has the highest degree of change and elevation has the lowest, showing that the most influential risk factors on exposure are land use, snail infection rate and distance to water bodies in that order, and the least influential factors are slope and elevation (Table 2).

Table 2 Sensitivity of exposure to risk factors using entropy reduction (variables are listed in order of influence on exposure)

Node	Degree of entropy reduction	% of influence to the network
PAS	0.07149	28.0
SI	0.06524	25.3
CC	0.04708	18.3
LU	0.04138	16.0
DWB	0.02868	11.1
SLP	0.00291	1.1
E	0.00066	0.2

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 11 of 15

Our findings show that approximately 63% of the study area has high probability of exposure values (> 50%). This is mainly explained by the predominance of agricultural fields in the area (Fig. 7b) and the distance to water bodies results, which indicate that approximately 80% of the urban areas can access water bodies following routes < 500 m. Lowest and highest distance values between urban areas and water bodies are 7.6 m and 5.7 km, respectively, with a mean of 1.4 km (Fig. 8).

Validation

For the first validation, the results show an increase in the probability of exposure as the proportion of human cases also increases, except for 17% of human cases where a reduction in the probability of exposure of 35.8% can be observed (Table 3). For the second validation, four groups of buffers were observed: Group A with one positive case, Group B with two positive cases and Groups C and D with four and five positive cases, respectively (Fig. 9). A low correlation was found between probability of exposure and percentage of human cases within the groups (linear correlation, $R^2 = 0.3$). For the first three groups (A, B and C) the probability of exposure increases while the percentage of human cases also increases. For Group D, the group with more positive cases, a minor decrease in the probability of exposure can be observed (Fig. 10). This could be explained by the distance to water bodies that has a negative correlation (Group C: -0.3, Group D: -0.02) with the

probability of exposure values (0.47–0.55) calculated from our sBN for groups C ($R^2 = 0.98$) and D ($R^2 = 0.96$) (Fig. 11). For instance, for Groups A and B with one and two positive cases respectively, the distance to water bodies is higher for Group A (~980 m) than for Group B (~177 m), with an average exposure value of approximately 0.47 and 0.48, respectively (Fig. 10). Likewise, for Groups C and D, the distance to water bodies is higher for Group D (~1100 m) than for Group C (~490 m), with an average probability of exposure values equal to 0.55 and 0.49, respectively (Fig. 10).

Discussion

Several studies have modelled snail distribution as input information for risk prediction of schistosomiasis [2, 20, 22, 47, 48, 51], in order to guide prevention (sanitary and hygiene conditions of the population) and control (mass drug administration campaigns in the community) strategies for schistosomiasis infection. These approaches are inadequate spatial decision support tools since they have not accounted for snails' infection status or people's exposure to infection (i.e. contact of people with snails' sites). In this study we demonstrate a novel approach to delineate spatial areas of exposure to *S. japonicum* infection by accounting for the distribution of infected and non-infected snails, and considering the human interaction with active transmission sites. This was done by accounting for the cost of the community

Fig. 8 Nearest route calculation from urban points to water bodies and DWB ordinary kriging interpolation
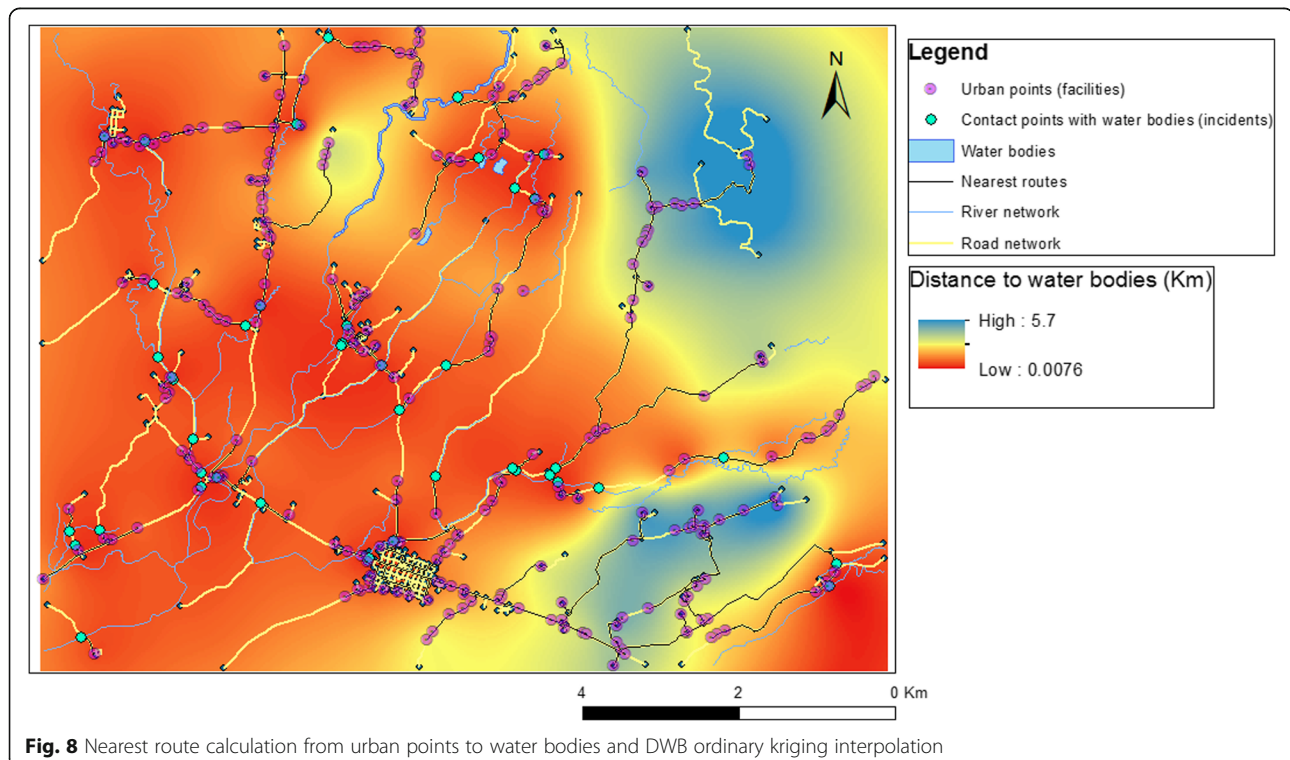
Table 3 Percentage of human cases falling within probabilities of high exposure values

No. of human cases	% of human cases	Probability of exposure
1	8.3	41.2
2	16.7	35.8
3	25.0	50.8
6	50.0	55.6

to access water bodies and potential sites where snails may be present.

Our results suggest that the predominance of high probabilities of exposure values (> 50%) in the study area are explained by the presence of wet soils and agriculture land in the zone, but also by the distance from urban areas to nearby water bodies (< 5 km). This was expected given that land use is a highly influencing risk factor on exposure after potential accessible sites (Table 2), and also because of the initial high weights given to *LU* and *DWB* (Table 1).

Our results demonstrate that for short distances to water bodies, the probability of a community to be exposed to *S. japonicum* is high (Fig. 8). This was explained by the probability of exposure map and the relative influence of *DWB* on exposure. Although *DWB* is the fifth influencing factor on exposure (Table 2), it is the only influencing factor on community cost, which is the third most important variable of the network (Table 2). Based on our results we propose that future

studies utilise the nearest distance to water bodies following a road instead of the commonly used Euclidean distance [51–53], since the former provides a more accurate representation of community access to water bodies, as it accounts for the nearest path from human dwellings to potential infection foci.

We postulated that the proportion of human *S. japonicum* cases was higher in areas predicted to have a higher probability of exposure. Our validation procedure using overlaying proportions in the four groups of buffers surrounding nearby *S. japonicum* cases, demonstrated a positive correlation for three groups. Although the number of validation points is somewhat low for a total validation, overlying proportions of exposure to schistosomiasis infection suggest a correlation between potential areas of exposure and the disease in the presence of limited survey data.

Utility of modelling the geographical probability of *S. japonicum* exposure

Modelled schistosomiasis exposure areas account for the transmission processes occurring between the environment containing infective stages of *S. japonicum* or intermediary hosts (snails), and the susceptible hosts (humans and livestock). From a public health perspective, the provision of maps that define the geographical limits of probability of exposure to *S. japonicum* infected areas could help target local schistosomiasis control strategies to communities more likely to contact contaminated

Fig. 9 Buffers around surveyed human cases points. Letters show the grouped buffers based on points location

Araujo Navas et al. Parasites & Vectors (2018) 11:465

Page 13 of 15
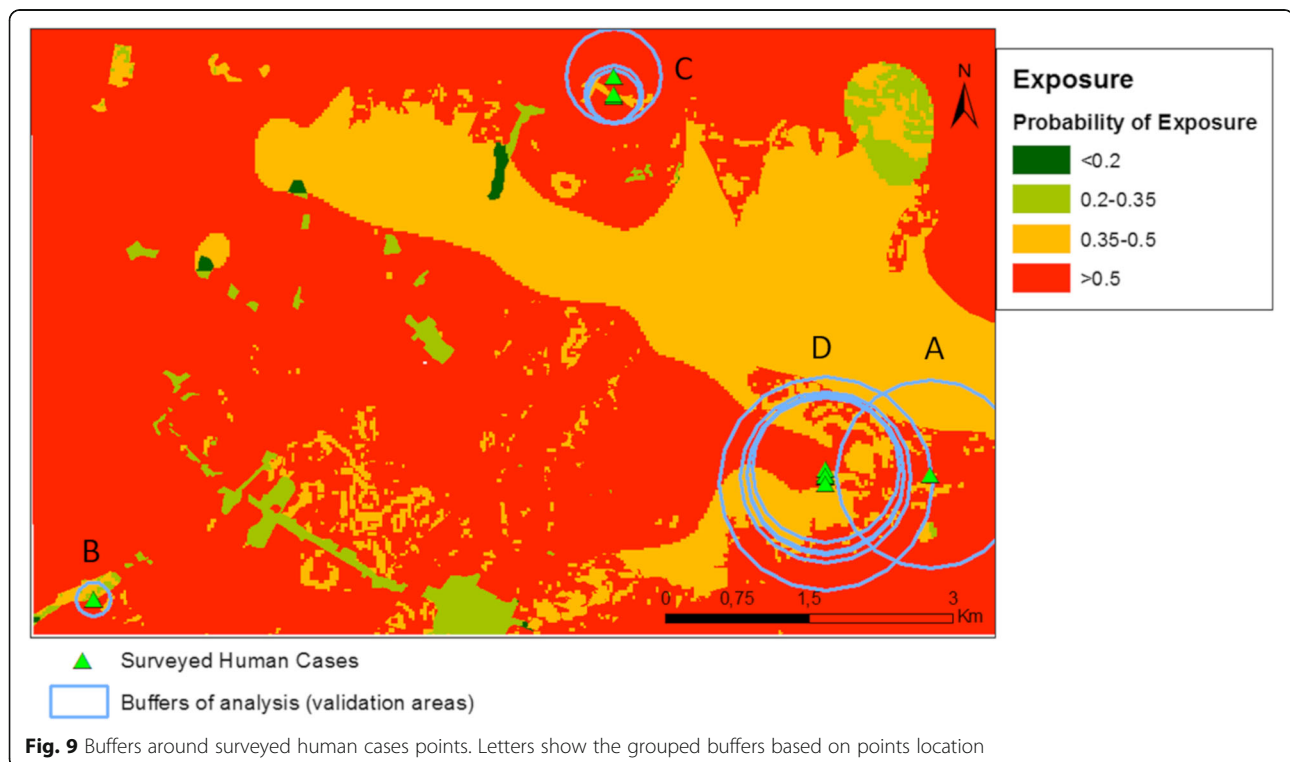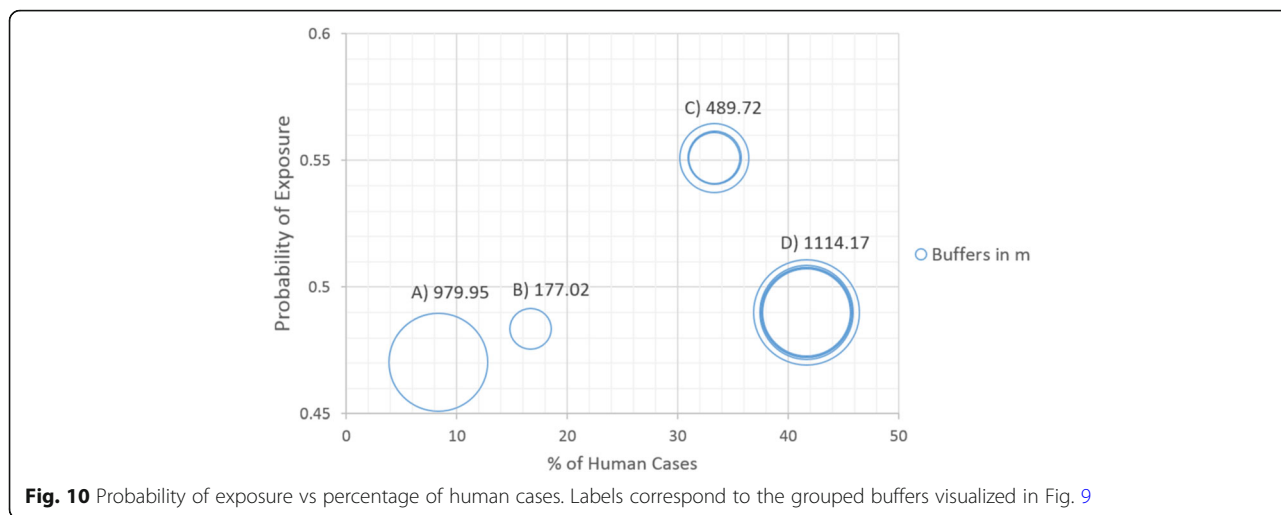
Fig. 10 Probability of exposure vs percentage of human cases. Labels correspond to the grouped buffers visualized in Fig. 9

environments and thereby improve the efficiency of mass drug administration campaigns. From a spatial modelling perspective, the availability of a predictive exposure map could serve as an important base map to obtain covariate values. By relating them to indicators of disease, we could possibly account for the positional mismatch between epidemiological survey data and environmental covariates, and improve the statistical modelling of *S. japonicum* infection.

Limitations of the study

A number of limitations should be accounted for in the interpretation of our results. Firstly, estimates of the

probability of exposure are highly influenced by the availability of snail infection estimates (Table 2). Due to the localized nature of the study, it was difficult to generate an adequate surface map that could properly explain snail infection distribution, constraining this map into a binary output with low and medium risk values (Fig. 7f). This might have an impact on the results and could be further improved by an increase of the study extent, and the number of survey points. In addition, whenever these data or new knowledge becomes available, the sBN developed in this study will enable a "rapid delineation" of potential exposure areas of *S. japonicum* by facilitating a flexible integration of exposure data as risk factors, and prior information derived from literature or expert knowledge [54].

Secondly, model validation procedures could be improved by including positive and negative human cases. Collecting data on livestock infection [23, 24, 30, 31] could also serve for validation as livestock infection, particularly carabao, has been suggested to play an important role in the transmission of *S. japonicum* in the Philippines [55].

Conclusions

In conclusion, the present study describes the nature of the positional exposure mismatch in the modelling of *S. japonicum* infection. Results of the present study suggest that the best way to address this mismatch should include the extraction of covariate values from potential exposure areas. A probabilistic method to delineate exposure areas in the absence of sufficient empirical survey data is proposed. Unlike other studies, the present sBN is adequate to delineate exposure areas based upon the contact of communities to water bodies and other potential sites of infection. We conclude that even with limited disease
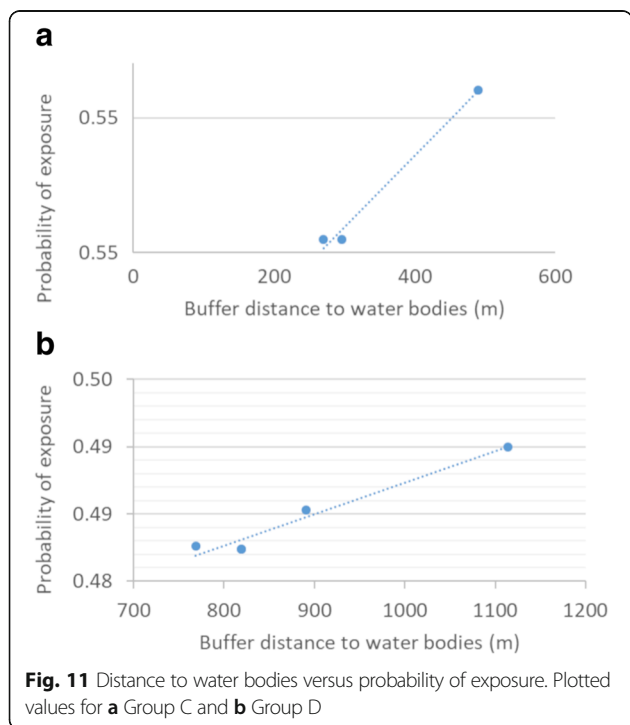
Fig. 11 Distance to water bodies versus probability of exposure. Plotted values for **a** Group C and **b** Group D

Araujo Navas et al. Parasites & Vectors (2018) 11:465

Page 14 of 15

survey data, it is possible to define potential exposure areas for schistosomiasis. Modelled exposure areas might be used to correct for positional mismatches and significantly improve disease predictions to better guide control programs to prevent and control schistosomiasis and other water-borne infections.

Additional files

Additional file 1: Table S1. Saaty's pairwise comparison matrix for Land Use. **Table S2.** Saaty's pairwise comparison matrix for Elevation. **Table S3.** Saaty's pairwise comparison matrix for Slope. **Table S4.** Saaty's pairwise comparison matrix for Distance to water bodies. **Table S5.** Saaty's pairwise comparison matrix for Snail infection rate. **Table S6.** Saaty's pairwise comparison matrix for all risk factors. **Table S7.** Total weights for all risk factors. Saaty's pairwise comparison tables for the risk factors and their categories. This file also includes the calculation of consistency indexes and ratios. (XLSX 31 kb)

Additional file 2: Table S8. Conditional probabilities for Land Use. **Table S9.** Conditional probabilities for Community cost. **Table S10.** Conditional probabilities for Exposure. Node probability tables. Conditional probability tables used for each one of the latent variable nodes: exposure (EX), potential accessible sites (PAS) and community cost (CC). (XLSX 16 kb)

Additional file 3: Table S11. Abbreviations used in the manuscript and variable notations. (XLSX 14 kb)

Abbreviations
BN: Bayesian network; DAG: Directed acyclic graph; GIS: Geographical information systems; GPS: Global positioning systems; NPT: Node probability tables; sBN: Spatial Bayesian network; SCH: Schistosomiasis

Availability of data and materials
The data supporting the conclusions of this article are included within the article and its additional files. The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Authors' contributions
ALAN, FO and RJSM contributed to conceptualization and study design. RJCF and LRL performed the field work and collected the data. ALAN analysed and interpreted the data. All authors were involved in drafting and revising the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
The human survey protocol was reviewed and approved by the University of the Philippines Manila Research Ethics Board (UPMREB) granted with ethical approval code UPM REB Code 2011-098. Written informed consent for human cases was obtained from participants. For children under the age of 16, informed consent was obtained from their parents or legal guardians.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. [2]UQ Spatial Epidemiology Laboratory, School of Veterinary Science, The University of Queensland, QLD, Gatton 4343, Australia. [3]Child Health and Environment Program, Child Health Research Centre, The University of Queensland, QLD, South Brisbane 4101, Australia. [4]Institute of Biology, College of Science, University of the Philippines Diliman, 1101 Quezon, Philippines. [5]Department of Parasitology, College of Public Health, University of the Philippines Manila, 1000 Manila, Philippines.

References
1. King CH, Dickman K, Tisch DJ. Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. Lancet. 2005;365:1561–9.
2. Walz Y, Wegmann M, Dech S, Vounatsou P, Poda J-N, N'Goran EK, et al. Modeling and validation of environmental suitability for schistosomiasis transmission using remote sensing. PLoS Neglect Trop D. 2015;9:e0004217.
3. Hotez PJ, Alvarado M, Basanez MG, Bolliger I, Bourne R, Boussinesq M, et al. The global burden of disease study 2010: interpretation and implications for the neglected tropical diseases. PLoS Neglect Trop D. 2014;8:9.
4. Araujo Navas AL, Hamm NAS, Soares Magalhães RJ, Stein A. Mapping soil transmitted helminths and schistosomiasis under uncertainty: a systematic review and critical appraisal of evidence. PLoS Neglect Trop D. 2016;10: e0005208.
5. Leenstra T, Acosta LP, Langdon GC, Manalo DL, Su L, Olveda RM, et al. Schistosomiasis japonica, anemia, and iron status in children, adolescents, and young adults in Leyte, Philippines. Am J Clin Nutr. 2006;83:371–9.
6. Coutinho HM, McGarvey ST, Acosta LP, Manalo DL, Langdon GC, Leenstra T, 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;192:528–36.
7. Jia TW, Zhou XN, Wang XH, Utzinger J, Steinmann P, Wu XH. Assessment of the age-specific disability weight of chronic schistosomiasis japonica. Bull World Health Organ. 2007;85:458–65.
8. Tarafder MR, Balolong E, Carabin H, Belisle P, Tallo V, Joseph L, 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.
9. Yang K, Wang XH, Yang GJ, Wu XH, Qi YL, Li HJ, Zhou XN. 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;38:1007–16.
10. Soares Magalhães RJ, Salamat MS, Leonardo L, Gray DJ, Carabin H, Halton K, et al. Geographical distribution of human Schistosoma japonicum infection in The Philippines: tools to support disease control and further elimination. Int J Parasitol. 2014;44:977–84.
11. 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;13:400–3.
12. Hay SI, 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;18:2899–930.
13. 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.
14. Hamm NAS, Soares Magalhães RJ, Clements ACA. Earth observation, spatial data quality, and neglected tropical diseases. PLoS Neglect Trop D. 2015;9: e0004164.
15. Zhang ZJ, Manjourides J, Cohen T, Hu Y, Jiang QW. Spatial measurement errors in the field of spatial epidemiology. Int J Health Geogr. 2016;15:12.
16. Soares Magalhães RJ, Clements ACA, Patil AP, Gething PW, Brooker S. The applications of model-based geostatistics in helminth epidemiology and control. Adv Parasitol. 2011;74:267–96.

Araujo Navas *et al. Parasites & Vectors* (2018) 11:465

Page 15 of 15

17. Cadavid Restrepo AM, Yang YR, McManus DP, Gray DJ, Giraudoux P, Barnes TS, et al. The landscape epidemiology of echinococcoses. Infect Dis Poverty. 2016;5:13.

18. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI, Gething PW. Re-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach. Malar J. 2015;14:68.

19. Moodley I, Kleinschmidt I, Sharp B, Craig M, Appleton C. Temperature-suitability maps for schistosomiasis in South Africa. Ann Trop Med Parasitol. 2003;97:617–27.

20. Stensgaard AS, Jorgensen A, Kabatereine NB, Rahbek C, Kristensen TK. Modeling freshwater snail habitat suitability and areas of potential snail-borne disease transmission in Uganda. Geospat Health. 2006;1:93–104.

21. Stensgaard AS, Utzinger J, Vounatsou P, Hurlimann E, Schur N, Saarnak CFL, et al. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? Acta Trop. 2013;128:378–90.

22. Guo JG, Vounatsou P, Cao CL, Utzinger J, Zhu HQ, Anderegg D, et al. A geographic information and remote sensing based model for prediction of *Oncomelania hupensis* habitats in the Poyang Lake area, China. Acta Trop. 2005;96:213–22.

23. Yang GJ, Vounatsou P, Tanner M, Zhou XN, Utzinger J. Remote sensing for predicting potential habitats of *Oncomelania hupensis* in Hongze, Baima and Gaoyou lakes in Jiangsu Province, China. Geospat Health. 2006;1:85–92.

24. Zhang ZJ, Bergquist R, Chen DM, Yao BD, Wang ZL, Gao J, Jiang QW. Identification of parasite-host habitats in Anxiang County, Hunan Province, China, based on multi-temporal China-Brazil Earth Resources Satellite (CBERS) Images. PLoS One. 2013;8:9.

25. Leonardo L, Rivera P, Saniel O, Solon JA, Chigusa Y, Villacorte E, et al. New endemic foci of schistosomiasis infections in the Philippines. Acta Trop. 2015;141:354–60.

26. Leonardo L, Acosta LP, Olveda RM, Aligui GDL. Difficulties and strategies in the control of schistosomiasis in the Philippines. Acta Trop. 2002;82:295–9.

27. Zhou XN, Bergquist R, Leonardo L, Yang GJ, Yang K, Sudomo M, Olveda R. Schistosomiasis japonica: control and research needs. Adv Parasitol. 2010;72: 145–78.

28. Olveda RM, Tallo V, Olveda DU, Inobaya MT, Chau TN, Ross AG. National survey data for zoonotic schistosomiasis in the Philippines grossly underestimates the true burden of disease within endemic zones: implications for future control. Int J Infect Dis. 2016;45:13–7.

29. Liu Z, Li C, Tang L, Zhou X, Ma L, Liu C. Prediction of *Oncomelania hupensis* (vector of schistosomiasis) distribution based on remote sensing data and fuzzy information theory. In: Geoscience and Remote Sensing Symposium (IGARSS). 26–31 July 2015, Milan, Italy. Milan: IEEE International; 2015. p. 4408–11.

30. Gao FH, Abe EM, Li SZ, Zhang LJ, He JC, Zhang SQ, et al. Fine scale spatial-temporal cluster analysis for the infection risk of schistosomiasis japonica using space-time scan statistics. Parasit Vectors. 2014;7:578.

31. Yang K, Li W, Sun LP, Huang YX, Zhang JF, Wu F, Hang DR, et al. Spatio-temporal analysis to identify determinants of *Oncomelania hupensis* infection with *Schistosoma japonicum* in Jiangsu Province, China. Parasit Vectors. 2013;6:138.

32. Pesigan TP, Hairston NG, Jauregui JJ, Garcia EG, Santos AT, Santos BC, Besa AA. Studies on *Schistosoma japonicum* infection in the Philippines 2. The molluscan host. Bull World Health Organ. 1958;18:481–578.

33. Geological Survey US. Global Data Explorer. 2017. https://gdex.cr.usgs.gov/gdex/. Accessed 1 Aug 2017.

34. Project OSM. Planet OSM. 2017. https://planet.osm.org. Accessed 21 Nov 2017.

35. Schultz M, Voss J, Auer M, Carter S, Zipf A. Open land cover from OpenStreetMap and remote sensing. Int J Appl Earth Obs. 2017;63:206–13.

36. Fonte CC, Minghini M, Patriarca J, Antoniou V, See L, Skopeliti A. Generating up-to-date and detailed land use and land cover maps using OpenStreetMap and GlobeLand30. ISPRS Int J Geo-inf. 2017;6:125.

37. Chen J, Chen J, Liao A, Cao X, Chen L, Chen X, et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J Photogramm. 2015;103:7–27.

38. Project OSGF. QGIS, a free and open source geographic information system. 2018. https://www.qgis.org/en/site/. Accessed 29 Nov 2017.

39. ESRI. ArcGIS Desktop: Release 10. 2011. http://www.esri.com/news/releases/10_2qtr/arcgis10-download.html. Accessed 15 Dec 2017.

40. Pebesma E, Graeler B. Spatial and spatio-temporal geostatistical modelling, prediction, package 'gstat'. In: The Comprehensive R Archive Network: R; 2017.

41. Corporation NS. NeticaTM application for belief networks and influence diagrams: User's guide. Vancouver: Norsys Softwate Corporation; 1998.

42. Bottcher SG, Dethlefsen C. deal: a package for learning Bayesian networks. J Stat Softw. 2003;8:20.

43. Bishop CM. Pattern recognition and machine learning. New York: Springer Science and Business Media; 2006.

44. Nielsen TD, Jensen FV. Bayesian networks and decision graphs. New York: Springer Science and Business Media; 2009.

45. Fenton N, Neil M. Risk assessment and decision analysis with Bayesian networks. Boca Raton: CRC Press; 2012.

46. Saaty TL. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors. The analytic hierarchy/network process. Rev Real Acad Cienc. 2008;102:251–318.

47. Hu Y, Xia CC, Li SZ, Ward MP, Luo C, Gao FH, et al. Assessing environmental factors associated with regional schistosomiasis prevalence in Anhui Province, Peoples' Republic of China using a geographical detector method. Infect Dis Poverty. 2017;6:8.

48. Zhang ZJ, Carpenter TE, Lynn HS, Chen Y, Bivand R, Clark AB, et al. Location of active transmission sites of *Schistosoma japonicum* in lake and marshland regions in China. Parasitology. 2009;136:737–46.

49. Ajakaye OG, Adedeji OI, Ajayi PO. Modeling the risk of transmission of schistosomiasis in Akure North Local Government Area of Ondo State, Nigeria using satellite derived environmental data. PLoS Neglect Trop D. 2017;11:e0005733.

50. Marcot BG, Steventon JD, Sutherland GD, McCann RK. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Can J Forest Res. 2006;36:3063–74.

51. Zhu HR, Liu L, Zhou XN, Yang GJ. Ecological model to predict potential habitats of *Oncomelania hupensis*, the intermediate host of *Schistosoma japonicum* in the mountainous regions. China. PLoS Neglect Trop Dis. 2015;9:e0004028.

52. Clements ACA, Lwambo NJS, Blair L, Nyandindi U, Kaatano G, Kinung'hi S, et al. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. Trop Med Int Health. 2006;11:490–503.

53. Kabatereine NB, Brooker S, Tukahebwa EM, Kazibwe F, Onapa AW. Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. Trop Med Int Health. 2004;9:372–80.

54. Smith CS, Howes AL, Price B, McAlpine CA. Using a Bayesian belief network to predict suitable habitat of an endangered mammal - the Julia Creek dunnart (*Sminthopsis douglasi*). Biol Conserv. 2007;139:333–47.

55. Gordon CA, Acosta LP, Gray DJ, Olveda RM, Jarilla B, Gobert GN, et al. High prevalence of *Schistosoma japonicum* infection in Carabao from Samar Province, the Philippines: implications for transmission and control. PLoS Neglect Trop D. 2012;6:7.

56. Steinmann P, Zhou XN, Li YL, Li HJ, Chen SR, Yang Z, et al. Helminth infections and risk factor analysis among residents in Eryuan county, Yunnan Province, China. Acta Trop. 2007;104:38–51.

57. Head JR, Chang H, Li QN, Hoover CM, Wilke T, Clewing C, et al. Genetic evidence of contemporary dispersal of the intermediate snail host of *Schistosoma japonicum*: movement of an NTD host is facilitated by land use and landscape connectivity. PLoS Neglect Trop Dis. 2016;10:e0005151.

58. Kloos H, Gazzinelli A, Van Zuyle P. Microgeographical patterns of schistosomiasis and water contact behavior; Examples from Africa and Brazil. Mem Inst Oswaldo Cruz. 1998;93:37–50.