Geospatial Health 6(1), 2011, pp. 41-51

Spatial patterns and eco-epidemiological systems – part I: multi-scale spatial modelling of the occurrence of Chagas disease insect vectors

Emmanuel Roux[1], Annamaria de Fátima Venâncio[2], Jean-François Girres[3], Christine A. Romaña[4]

[1]ESPACE-DEV, UMR228 IRD/UMII/UR/UAG, Institut de Recherche pour le Développement, Cayenne, French Guiana; [2]Instituto do Meio Ambiente e Recursos Hídricos do Estado da Bahia e Centro de Desenvolvimento Sustentável da Universidade de Brasília, Brazil; [3]IGN, COGIT, Saint Mandé, France/IRD, ESPACE-DEV (UMR228), Cayenne, French Guiana; [4]Université Paris Descartes, Paris, France/IRD, ESPACE-DEV (UMR228), France

Abstract. Studies that explicitly and specifically take into account the spatial dimension within the study of eco-epidemiological systems remain rare. Our approach of modelling the spatial and/or temporal properties of the entomological and/or epidemiological data before their mapping with possible explanatory variables, objectively underline the significant patterns at different scales. The domiciliary and peri-domiciliary presence and abundance of juvenile and adult vectors of the Chagas disease (*Triatoma sordida* and *Panstrongylus geniculatus*) in Bahia state in northeast Brazil, has been modelled by automatically identifying significant multi-scale spatial patterns of the entomological data by the application and adaptation of the spatial modelling methodology proposed by Dray et al. (2006) and based on principal coordinate analysis of neighbour matrices. We found that entomological data can be modelled by a set of eigenvectors that present a significant Moran's *I* index of spatial autocorrelation. The models for juvenile and adult vectors are defined by 28 and 32 eigenvectors that explain 82.3% and 79.9%, respectively, of the total data variances. The results support insect presence as the outcome both of a local scale "near-to-near" dispersal and an infestation from the wild, surrounding environment that produces a higher insect density at the village periphery.

Keywords: principal coordinate analysis of neighbourhood matrices, spatial modelling, Chagas disease, vectors, Brazil.

Introduction

American trypanosomiasis or Chagas disease is a vector-borne disease, caused by *Trypanosoma cruzi* (Chagas, 1909) and transmitted by insects of the Triatominae subfamily. Environmental conditions, including human behaviour, dwellings and peri-domiciliary characteristics that can potentially influence vector presence, abundance and infestation, are not well known for all species. In fact, there are up to 45 triatomine species in Brazil, which all have different ecological niches (palm trees, rocks anfractuosities, animal burrows, etc.) and behaviour.

Environmental conditions partially determine the presence, density and spatio-temporal distribution of pathogens, vectors and hosts and thus the epidemics, the (re)-emergence and/or the endemicity of many vector-borne diseases. In this framework, much deals with mapping environmental and geo-localized epidemiological data (prevalence and, ideally, incidence), while transmission of the disease is considered a "black box" (Peterson, 2007). This approach is justified when the epidemiological data are aggregated making geo-localised cases, and associated transmission zones coincide spatially and temporally. However, this provides only a poor insight into how the eco-epidemiological system functions. In fact, understanding the true mechanisms of transmission demands a finer spatial and temporal resolution than is normally applied and, without a strong hypothesis on possible transmission locations, the vector and its ecology become the focus of the system (Romaña, 2004; Peterson, 2006, 2007). Such studies map environmental characterisation, notably derived from high, or very high, resolution image processing, geo-localised entomological data (presence/absence and/or abundance data) and, but rarely jointly, epidemiological data (Girod et

Corresponding author:
Emmanuel Roux
ESPACE-DEV, UMR228 IRD/UMII/UR/UAG
Institut de Recherche pour le Développement
Cayenne, French Guiana
Tel. +594 (0)5 94 29 92 77 Fax +594 (0)5 94 31 98 55
E-mail : emmanuel.roux@ird.fr

al., 2011). However, when using remotely sensed data, a compromise between spatial and temporal resolution has to be found (Kitron et al., 2006). Moreover, the availability of remotely sensed data can be particularly restricted in regions with persistent cloud cover, which is often the case in tropical zones. Beyond these practical limits, such an approach can also be questioned from a theoretical point of view. In fact, the spatially observable resultants of the eco-epidemiological system depend on environmental, social, demographic and behavioural characteristics as well as on the interactions between these factors at different scales. Thus, associations between the epidemiological and/or entomological data and the space dimension encompass the links between such data and the factors listed above. We suggest that a specific study of the spatial and/or temporal properties of the entomological and/or epidemiological data should be carried out before linking such data with any explanatory variables, as this approach would permit the exploration of the spatial and/or temporal properties at different scales.

Numerous epidemiological studies on Chagas disease take into account the spatial dimension in an implicit way by integrating environmental variables having by nature a strong spatial basis (e.g. Costa et al., 2002; Peterson et al., 2002; Dumonteil et al., 2004; Abad-Franch et al., 2010). However, studies that explicitly take into account the spatial dimension remain rare. For example, Ramirez-Sierra et al. (2010) studied domiciliary infestation of four villages on the Yucatan peninsula, Mexico with regard to insect sex, infection status and distance between houses and the village boundary, the latter variable being suggested by observations and the outcome of a model proposed by Slimi et al. (2009). In a comparable way, Roux et al. (2009) used several attributes to explicitly characterise the spatial structure of a Brazilian village in the semi-arid region of Bahia: (i) distance to the village centre; (ii) distance to the specific house where the majority of the insects was found; and (iii) relative situations of the domiciliary unit (central or peripheral and connected or isolated). Such a characterisation contributed to the explanation of the presence of two triatomines: Panstrongylus geniculatus (Latreille, 1811) (adults) and Triatoma sordida (Stål, 1859) (juveniles). These two characterisations of the village structure remain subjective. Among works that specifically took into account the spatial dimension, Khan et al. (2010) used both spatial analysis and remote sensing to study the infestation of Argentinean communities by Triatoma infestans (Klug, 1834). The Getis-Ord local spatial statistic was used to identify spatial clusters corresponding to "hotspots of high-prevalence domestic infestation", while multivariate regression analysis was applied to identify the environmental and demographic factors that could explain the inclusion of a community in a hotspot. To explain re-infestation, Kitron et al. (2006) used the Getis-Ord local spatial statistic as focal statistic and the local K-function in order to identify clusters around *a priori* known sources for *T. infestans* and *T. guasayana* (Wygodzinsky and Abalos, 1951) dispersal. In addition, Slimi et al. (2009) have proposed a cellular automata model that permits the explicit consideration of both temporal and spatial dimensions to simulate the infestation in a "virtual" village, while Barbu et al. (2010) put forward a spatially explicit model to reconstruct observed spatial and temporal patterns of house infestation by *T. dimidiata* in a village on the Yucatan peninsula, Mexico.

The objective of our approach, schematically presented in Fig. 1, was to investigate the application of a method for the identification and modelling of spatial multi-scale patterns (part I) and the application of a multivariate descriptive data analysis to explain the patterns identified (part II). The applicative context is the study of the presence and abundance of Chagas disease vectors (*T. sordida* and *P. geniculatus*) in a village in a semi-arid region of Bahia, Brazil. First, we apply the modelling approach proposed by Dray et al. (2006), based on the principal coordinates analysis of a neighbour matrix (PCNM) (Borcard et al., 2004), consisting in the modelling of the spatial structure of a set of sampling sites, by orthonormal eigenvectors that maximise the Moran's *I* index of spatial auto-correlation and thus characterise significant spatial patterns. Borcard et al. (2004), Dray et al. (2006) and Bellier et al. (2007) have investigated to what extent these eigenvectors (considered as variables) explain one (or a set of) response variable(s) and Dray et al. (2006) proposed a data-driven model selection method based on the Akaike information criterion (AIC). In this first part of our communication, the entomological data and the methodology are described in a specific applicative framework that emphasises the adaptations of Dray et al.'s work (2006).

Materials and methods

Study area

Santa Rita, belonging to the Iraquara municipality, is a rural locality at an altitude of about 700 m above

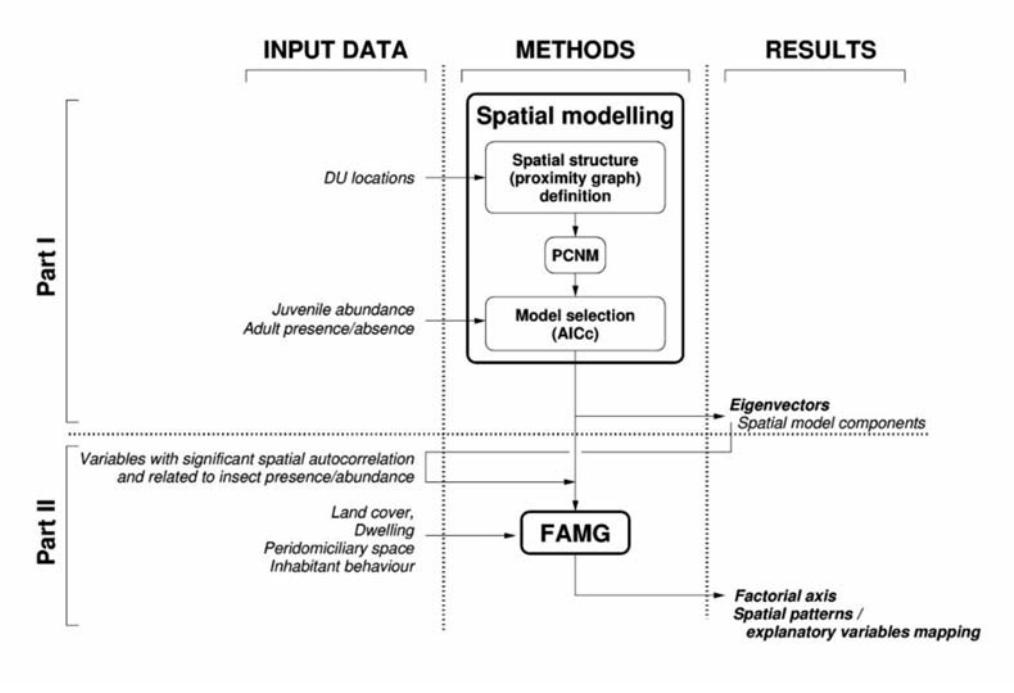

Fig. 1. Schematic representation of the overall methodology used. AICc: corrected Akaike information criterion; DU: domiciliary unit; FAMG: factorial analysis of mixed groups; PCNM: principal coordinate analysis of neighbour matrices.

sea level with about 500 inhabitants in the central part of Bahia, in northeast Brazil. It is part of the environmental protection area Marimbus/Iraquara. A subjective and broad-scale observation of the spatial structure of the village shows a densely inhabited zone (here referred to as the main hamlet) and four isolated groups of three to five houses, one in the south, denoted *S* in Fig. 2, and three in the north, denoted *N, NN1 and NN2*, respectively.

The main hamlet and surrounding areas is shown in Fig. 2, which is a SPOT-5 panchromatic image with a spatial resolution of 2.5 m emanating from the Seas Guyane Project (http://www.seas-guyane.org) and acquired, on 15 January 2007.

Collection methodology

The domiciliary unit (DU), i.e. the human dwelling and its peri-domiciliary space with all annexes as defined by Silveira and Rezende (1994), constituted the sampling unit. For each of the 132 village DUs, geo-localized by means of a global positioning system (GPS) instrument (GARMIN eTrex Legend®), eggs, juveniles, adults and exuvii were collected from October to December, 2007. In the homes, the insects were searched for in cracks and crevices of walls, under mattresses, behind cabinets and tables, in boxes or in piles of clothing or objects. In the peri-domiciliary spaces, every possible insect niche was considered:

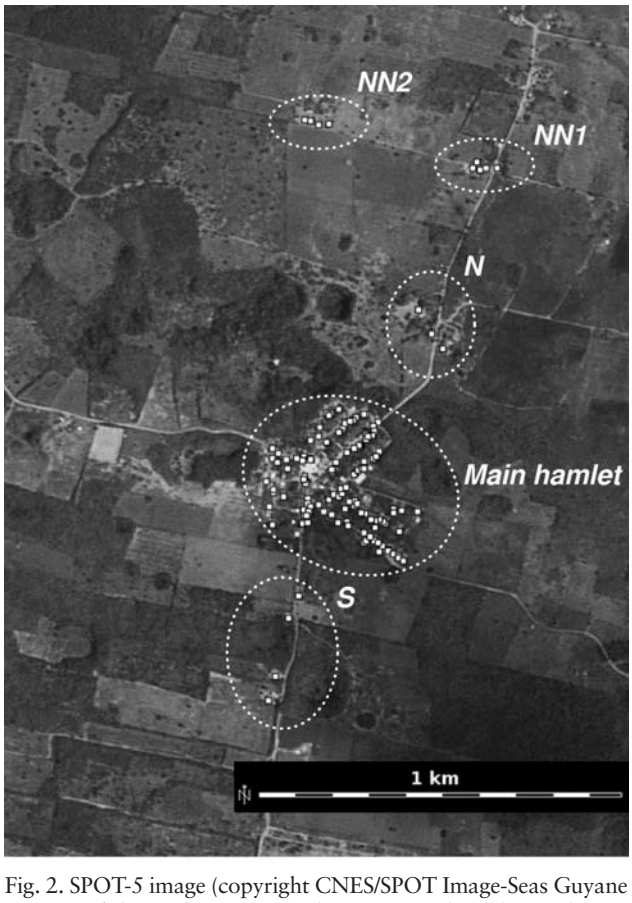

Fig. 2. SPOT-5 image (copyright CNES/SPOT Image-Seas Guyane Project) of the study site in panchromatic mode and spatial resolution of 2.5 m (acquisition date: 15 January 2007). White squares represent houses, geo-positioned by means of a GPS. Dotted lines delimit the groups of dwellings mentioned in the text.

plants (alive or dead), fences, henhouses, pig-sties corrals, warehouse deposits, stacks of tiles, bricks, stone slabs and wood.

Entomological data

Table 1 summarises the results of triatomine captures and Fig. 3 represents the data spatially. Overall 195 eggs and 53 exuvii were found, while 154 juveniles of the *T. sordida* species were captured. Twelve juveniles were not identified, but there is a strong presumption that these also belonged to the *T. sordida* species. Similarly, due to the almost systematic presence of *T. sordida* juveniles with eggs and exuvii, these two latter stages were also assumed to belong to the *T. sordida* species.

The adults found belonged to the species *T. sordida*, *T. lenti* (Sherlock and Serafim, 1967), *T. pseudomaculata* (Correa and Espinola, 1964), *P. lutzi* (Neiva and Pinto, 1923) and *P. geniculatus*. Forty-six *T. sordida* specimens were found but only five of the *P. geniculatus* species. In one unique DU, 35 *T. sordida* specimens were found, while the other infested DUs had only one individual (of variable species) each, except one where one *T. sordida* and one *P. geniculatus* were found. Adults and non-flying (here referred to as juveniles) specimens coexisted in five DUs, where four of them were exclusively infested by *T. sordida*, while one DU was infested by two *T. sordida* juveniles, one egg and one *T. pseudomaculata* adult.

In the analysis, we considered the cumulated number of eggs, exuvii and juveniles of the *T. sordida* species in order to characterise the non-flying stages (here referred to as juveniles). Presence of insects at these evolution stages indicates suitable conditions for reproduction and development. Given the very low number of individuals found for some of the species, we considered only two adult species in the analysis: *T. sordida* and *P. geniculatus* (Table 1).

Spatial modelling

We applied the spatial modelling method proposed by Dray et al. (2006) as it provides the most explanatory, spatial model given the entomological data. In the following, the different steps of the method are detailed in the specific framework of our application by focusing on the methodological adaptations. The implementation of the method was performed by means of the free and open-source geographical

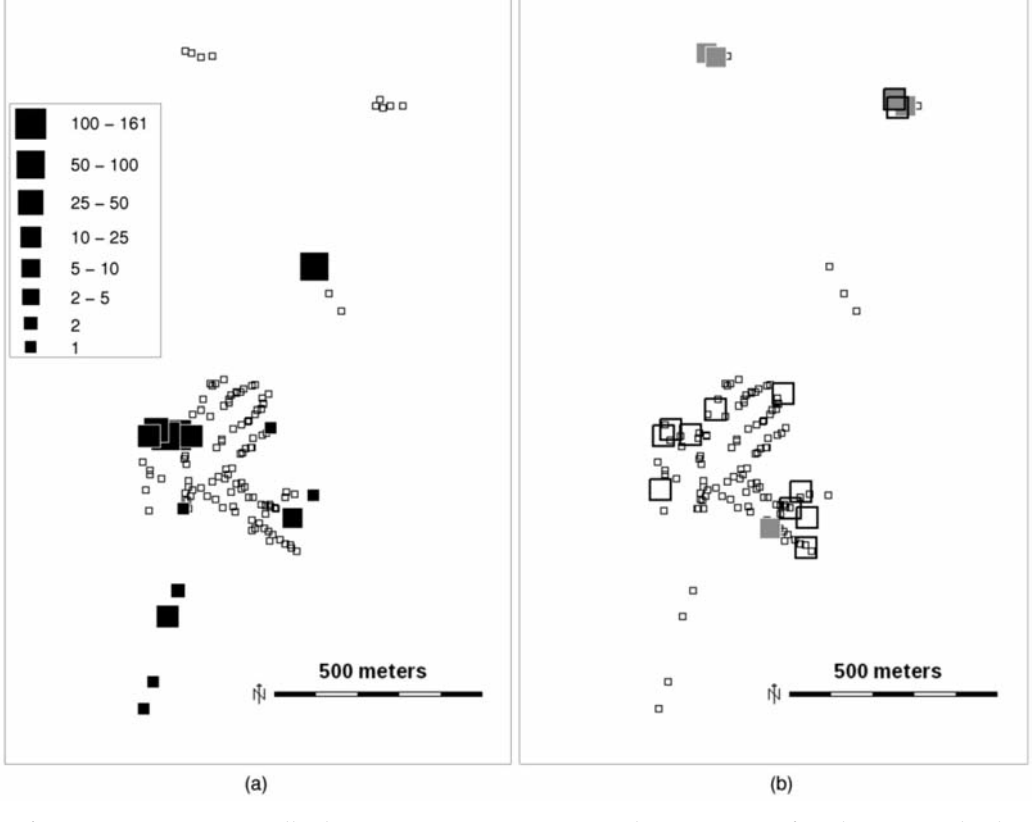

Fig. 3. Maps of vector occurrences. Small white squares represent DUs with no specimen found. (a) Juvenile abundance (black squares); (b) presence of adults of two species: *T. sordida* (large white squares) and *P. geniculatus* (grey squares).

Table 1. Summary of Triatomine capture results.

Development stage / species		Total found	Minimum found per DU	Maximum found per DU	Number of infested DUs
Exuvii, eggs and non-flying stages	Eggs	195	4	65	6
	Exuvii	53	1	23	10
	Juveniles (*Triatoma sordida*)	166	1	106	10
	Eggs + exuvii + juveniles (called juveniles in the study)	414	1	161	14
Adults	*Triatoma sordida*	46	1 (in all but one DU)	35 (in one DU)	12
	Panstrongylus geniculatus	5	1	1	5
	Panstrongylus lutzi	3	1	1	3
	Triatoma pseudomaculata	2	1	1	2
	Triatoma lenti	1	1	1	1

information system (GIS) GRASS (GRASS Development Team, 2010) and the free R environment for statistical computing (R Development Core Team, 2010).

Raw data pre-processing

The variable juveniles were first logarithmically transformed after the value of one had been added to the data, in order to make the distribution more symmetrical. No significant linear spatial trend was found for this variable. Consequently, the variable was not de-trended before being analysed (Borcard et al., 2004).

Due to the distribution of adults, we recoded the data to obtain information on presence and absence. We then computed the complete disjunctive table $Y=[y_{i,j}]_{i=1, ..., N;\ j=1, ...2\cdot S}$ with $y_{ij} \in \{0,1\}$, and finally performed the following transformation of the indicator variables as described by Legendre and Gallagher (2001) and Pagès (2002):

$$y'_{ij} = \sqrt{y_{..}} \cdot \frac{y_{ij}}{S \cdot \sqrt{y_{.j}}} = \sqrt{N} \cdot \frac{y_{ij}}{\sqrt{S} \cdot \sqrt{y_{.j}}} ,$$

where i and j represent the row and column indexes, respectively, $y_{..}$ the sum of all cells of the presence/absence species table, i.e. the product of the number of species (S) by the number of sites (N) and $y_{.j}$ the sum of the j^{th} column. This coding ensures that the χ^2 distance is used for site comparison in the following, which is consistent with a presence/absence coding of the species values and more generally with categorical variables. It provides other properties detailed below in terms of the spatial model selection criterion.

Neighbour matrices

Five types of neighbourhood structures (here also called graphs) were tested as mentioned by Dray et al. (2006): the Delaunay triangulation graph, the Gabriel graph, the relative neighbourhood graph, the minimum spanning tree, and distance based graph (dnn), considering that two sites i and j are neighbours if $d(i,j)<=\gamma$.

We explored several uniformly distributed values for γ and for each dataset (juveniles and adults). The minimum value was set to 75 m. Taking univariate and multivariate variograms into account (Wagner, 2003) for juvenile and adult data, respectively, the maximum value for γ was the maximum inter-site distance for which the empirical variogram was significant, i.e. 650 m and 775 m for juveniles and adults, respectively (Fig. 4). Consequently, the sets of values for juveniles and adults were S_γ^j = {75, 175, 275, 375, 475, 575, 650} and $S_\gamma^a = S_\gamma^j \cup \{675, 775\}$.

Neighbour matrice weighting

Previously presented spatial structures assign the same weight to all links (weight = 1). However, we may expect that DU similarity in terms of

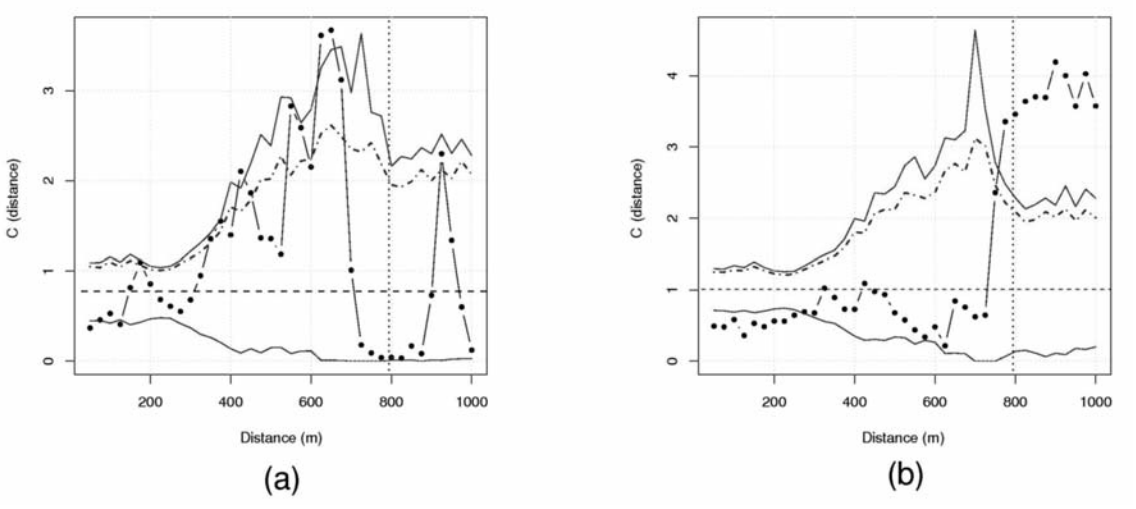

Fig. 4. Variograms of juveniles (a) and adults (b). Continuous lines represent the 2.5 and 97.5 percentiles estimated with 2,000 random permutations. Dashed and dotted lines represent the 95% percentile estimated with 2,000 random permutations. Dotted lines correspond to half the maximum distance between two sites.

insect/species presence and abundance is greater for close DUs than for distant ones. More specifically, in terms of dispersal behaviour, we may assume that insects have more chances to colonize proximate DUs than distant ones. Consequently, according to Dray et al. (2006), we defined different weighting functions as: $f_1=1-(d/dmax)^{\alpha}$ with $\alpha \in \{1, 2, 5, 10\}$ and $f_2=1/d^{\beta}$, with $\beta \in \{0.1, 0.2, 0.5, 1\}$.

Overall, 88 (11 graphs × 8 weighting functions) weighted graphs were tested for juveniles and 104 (13 graphs × 8 weighting functions) for adults.

Model generation and selection

The PCNM was computed for candidate spatial structures generated with the methodology detailed in Dray et al. (2006). For each eigenvector, the significance of Moran's *I* index value was tested with a 999 permutation procedure. Eigenvectors that presented significant spatial auto-correlation (P <0.01) in the sense of Moran's *I* were retained as proposed by Bellier et al. (2007). The eigenvectors were then sorted into descending order according to their capacity to explain the response variable (juvenile abundance) or the set of response variables (presence of adults), i.e. according to the explained variance provided by linear regression. They were then entered, one by one, for model definition (as many models as eigenvectors) and the corrected AIC (AICc) was computed to select the model with the minimum AICc value, i.e. the one that realised the best compromise between variance explanation and parsimony (for procedural details, see Dray et al., 2006).

Results

Juveniles

The minimum AICc value was obtained with the dnn graph with a distance γ equal to 575 m and the weighting function f_2 with $\beta = 0.2$ (Fig. 5). This model is defined by 28 eigenvectors and explains 82.3% of the total variance of juvenile abundance data (Fig. 6a). Fig. 7 shows the spatial representation of the first 12 eigenvectors that explain 66.9% of the total variance.

Figure 6b shows that the spatial model for juvenile data selected the majority of eigenvectors associated with positive values, i.e. with large-scale spatial patterns (Dray et al., 2006). However, it excludes the eigenvector associated with the maximum value. The first eigenvector of the model for juveniles emphasises

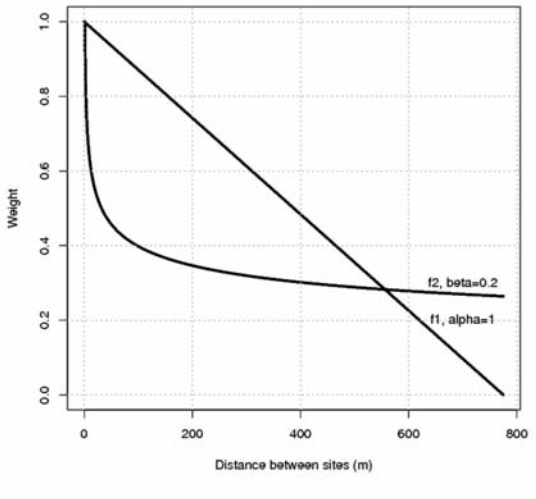

Fig. 5. Weighting functions selected for juveniles (f_2, $\beta = 0.2$) and adults (f_1, $\alpha = 1$).

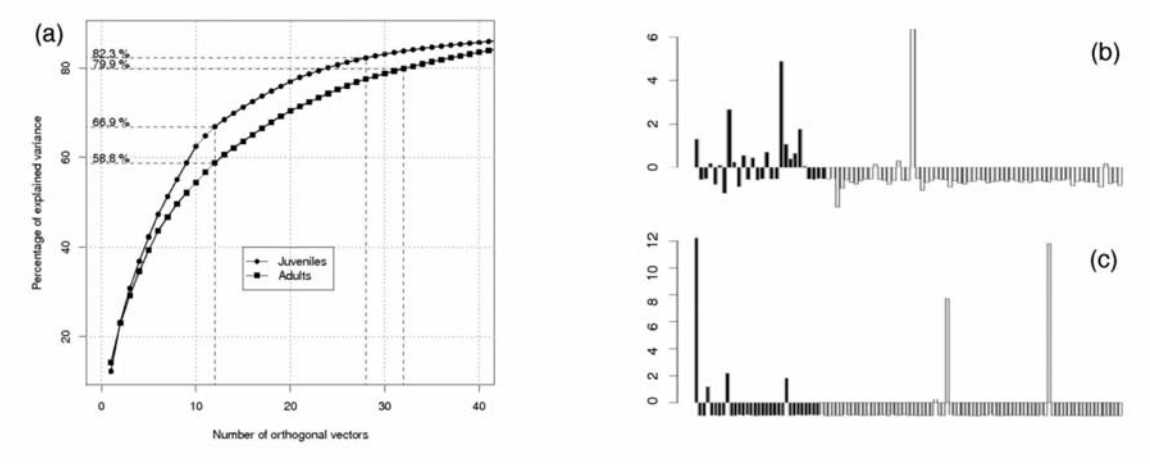

Fig. 6. (a) Cumulative percentage of the explained variances by the selected spatial models for juvenile and adult data. Cumulative percentages associated with the numbers of orthogonal eigenvectors chosen by the two models (28 and 32 eigenvectors for juveniles and adults, respectively) are shown, as well as the cumulative percentages associated with the first 12 eigenvectors depicted in Figs. 7 and 8. (b) and (c) Eigenvalues of the PCNM re-ordered by their capacity to explain juveniles and adults, respectively. Black bars correspond to the eigenvalues associated with selected eigenvectors by the AICc selection procedure.

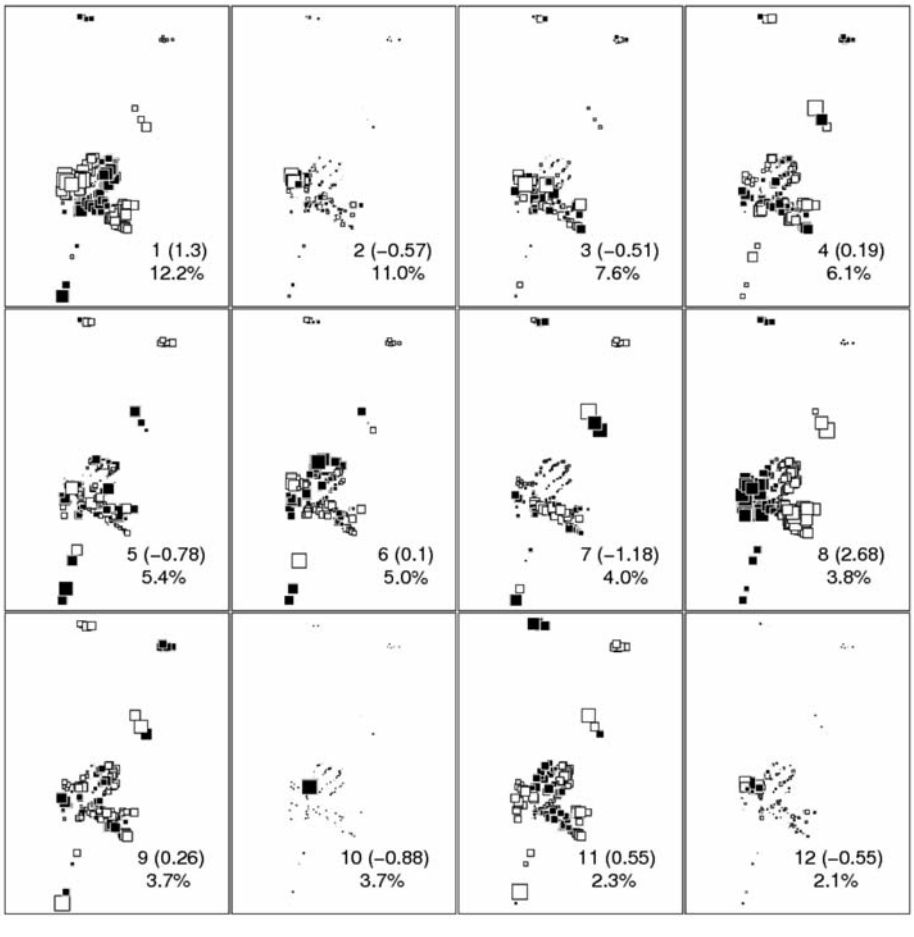

Fig. 7. Spatial representation of the first 12 eigenvectors of the selected spatial model for juveniles. Black and white squares correspond to positive and negative values of the eigenvector components, respectively. The square size is proportional to the absolute value of the eigenvector components. The eigenvectors are ranked by their capacity to explain juvenile abundance. This capacity is indicated by the percentage of variance of the juvenile data explained by the eigenvector. The other numbers correspond to the eigenvector number in the model (from 1 to 12) and to the associated eigenvalue. The eigenvalues are proportional to the Moran's *I* of spatial autocorrelation (Dray et al., 2006).

the DUs (black squares in Fig. 7) belonging to a band oriented southwest-northeast corresponding to the main road crossing the village, which separates the south-eastern from the north-western part. Eigenvectors associated with negative values were also selected, as they are associated with local-scale spatial patterns. In Fig. 7, spatial patterns associated with eigenvectors 2, 10 and 12 are shown to characterise very local features associated with a particular area of the village, near DU no. 117, where the great majority of the next-generation insects were found (43% of the exuvii, 16% of the eggs and 63% of the juveniles).

Adults

The minimum AICc value was obtained with the dnn graph with a distance γ equal to 775 m and the weighting function f_1 with $\alpha = 1$ (Fig. 5). It provides a linear weighting as a function of the distance between sites. The model is defined by 32 eigenvectors and explains 79.9% of the total variance of the presence of adults (Fig. 6a). Fig. 8 shows the spatial representation of the first 12 eigenvectors, which explains 58.8% of the total variance. The first eigenvector, corresponding to the highest value, clearly discriminates the isolated DU groups, especially the extreme northern ones denoted *NN1* and *NN2* in Fig. 2, associated with the presence of *P. geniculatus* (Fig. 3b). The fourth eigenvector is comparable with the first eigenvector of the juvenile model. The ninth eigenvector discriminates very central DUs (black squares) from the peripheral ones (white squares). Local patterns were associated with different parts of the main hamlet with a relatively high density of houses: the north-western part (eigenvector 2), the north-eastern part (eigenvectors 7 and 11), the south-eastern part (eigenvectors 3, 10 and 12) and the south-western part (eigenvector 8).

Discussion

Where insects are transiently present and/or in very low density in the domestic habitat, community participation has been found highly reliable for entomologic surveys and also shown to be more sensitive than

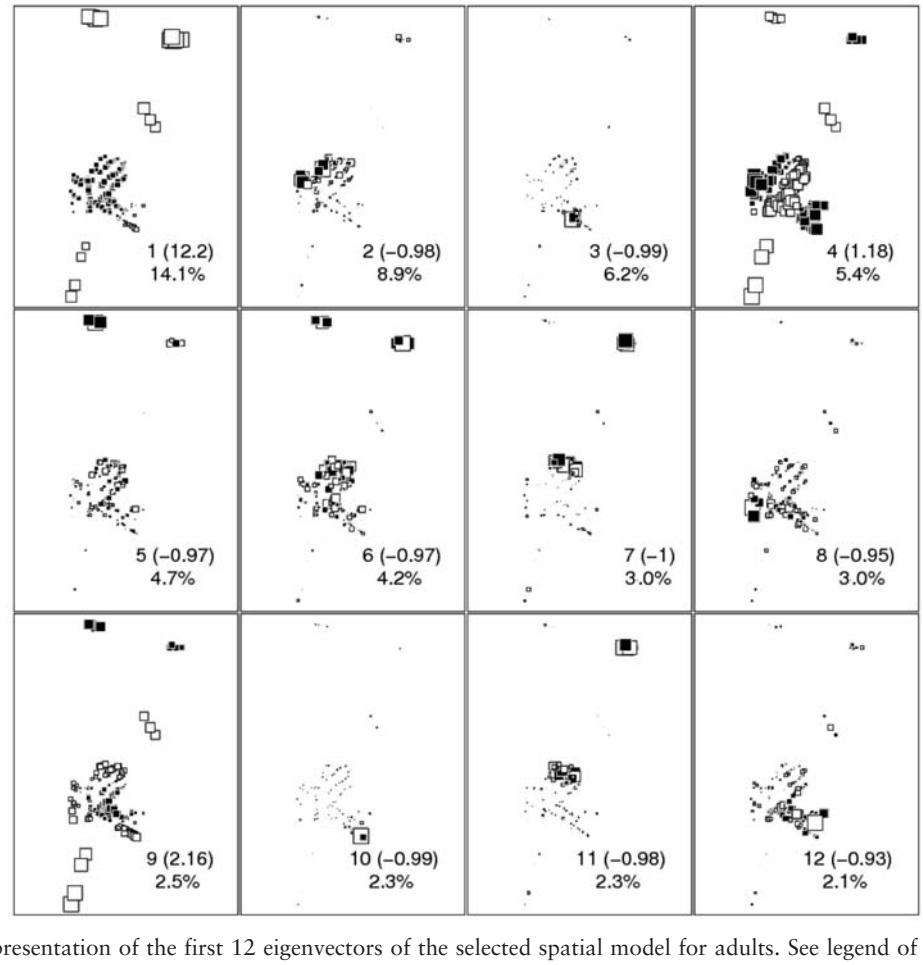

Fig. 8. Spatial representation of the first 12 eigenvectors of the selected spatial model for adults. See legend of Fig. 7 for explanations of symbols and numbers.

manual collection (Barbu et al., 2010; Ramirez-Sierra et al., 2010). Although we initially chose manual collection, it was performed by the same person to avoid operator bias. Moreover, this procedure permitted identification of the infestation process around highly infested sites, something that would not have been possible even by long-term, continuous collection based on community participation. Nevertheless, manual collection was necessary to make a first inventory of the situation, drawing attention to the public health issues associated with triatomines and to encourage people to bring insects to the Triatominae Information Centres (Postos de Informação de Triatomíneos = PIT), installed in the villages since the 1980s. However, it should be admitted that the chosen collection procedure prevents the study of temporal dynamics. To overcome this limitation, community participation should be seriously considered.

Dray et al. (2006) take into account all the eigenvectors provided by PCNM computation. This approach provides highly explanatory variables. However, such selected variables do not necessary present significant spatial patterns in the sense of Moran's *I*. The set of selected variables can comprise highly specific explanatory variables that can be difficult to interpret within the application context and that cannot be used for knowledge generalization.

In Bellier et al. (2007), only one spatial structure is defined corresponding to the truncated distance matrix proposed in Borcard et al. (2004). Only eigenvectors that present a significant spatial auto-correlation according to the Moran's *I* were selected. This results in a model with strong spatial correlations but a model that depends much on *a priori* and arbitrary chosen spatial structures. In this paper, we benefit from the advantages of these two approaches, while avoiding their drawbacks: (i) by performing a spatial model selection with no *a priori* spatial structure, and; (ii) by taking into account model components with a strong spatial correlation.

Given the data pre-processing used, model selection criteria differ for juveniles and adults, while being consistent with the data characteristics. For juveniles characterised by a quantitative variable, the selected spatial model realises the compromise between the model parsimony and the minimum variance of multiple linear regression residuals. For adults, on the other hand, investigating the presence/absence data given the weighting of the indicators of the complete disjunctive table, the selected model realises the compromise between the model parsimony and the maximum correlation ratio.

The first eigenvector of the spatial model for adult insects essentially explains the *P. geniculatus* species distribution. It corresponds to the first eigenvector of the PCNM, associated with a large-scale variation, more specifically with the distance to the village centre in the latitudinal direction. The application of the methodology for explaining spatial repartition of the *P. geniculatus* species is questionable due to the low number of infested sites (n = 5) and because this species was clearly associated with isolated DUs located in the northern part of the village, except for one specimen. A de-trended version of this species abundance data, i.e. the residuals of a multiple linear regression model on geographic coordinates, could have been considered (Borcard et al., 2004; Dray et al., 2006). However, due to the low number of infested sites, and the fact that only one individual has been found at each site, these de-trended data differed too much from the true observations. Moreover, we applied the methodology on the de-trended version of the adult data (Legendre and Gallagher, 2001). The selected model did not provide a better explanation and was not more parsimonious than the model selected with the proposed methodology.

Models obtained by the methodology used result in high numbers of components (28 and 32 for juveniles and adults, respectively) and interpreting each of them is a difficult task. As in Bellier et al. (2007), eigenvector clustering could result in a better insight regarding the different significant scales identified by the model. However, mapping of eigenvectors with explanatory variables (environment, behavioural, etc.), as done in part II of this paper, provides more information.

For both juvenile and adults, spatial patterns are associated with both local and global scales. However, the spatial model for juveniles is defined by a more localized neighbourhood (γ = 575 m) than the model for adults (γ = 775 m). Moreover, the weights of the links for the juvenile spatial structure decrease fast as function of the inter-site distance, whereas these weights decrease linearly with the inter-site distance for adults (Fig. 5). These results indicate that the spatial correlation for juveniles are associated with more local scales and can be interpreted as "near-to-near" dispersal. Barbu et al. (2010) have demonstrated that *T. dimidiata* disperse at rather small distances too. In our case, local-scale phenomena for juveniles seem essentially derived from a "near-to-near" dispersal of the insects associated with DU no. 117 that was host to the great majority of the insects (43% of the exuvii, 16% of the eggs and 63% of the juveniles). A similar infestation, or re-infestation, process around highly

infested sites has previously been observed for *T. infestans* and *T. guayasana* (Kitron et al., 2006; Ramirez-Sierra et al., 2010).

Large-scale spatial patterns appear to be linked to the distance to the centre of the main hamlet, this distance being positively correlated with the presence of insects. This suggests a colonisation from the wild environment surrounding the village and a progressive infestation towards the village centre as already observed and modelled (Slimi et al., 2009; Ramirez-Sierra et al., 2010; Barbu et al., 2010). Thus, by objective and quantitative approaches, two dispersal modalities can be derived: (i) infestation steming from the "natural" environment surrounding villages, and; (ii) local dispersal from one domiciliary unit of a village to the nearest one.

Conclusion

The methodology presented here makes it possible to spatially model the presence and abundance of disease vectors at different scales. It also supports the construction of a hypothesis regarding the dispersal behaviour of the insects. Infestation can originate from the "natural" environment surrounding villages as well as be the result of local dispersal from one domiciliary unit of a village to another.

The methodology presented here can be applied to other geo-localised data, e.g. disease cases. Naturally, many factors influence the eco-epidemiological system and consequently the spatial patterns of its observable characteristics (presence and abundance of the vectors/hosts, prevalence/incidence of disease). In Part II of this paper, we propose a method to hierarchically explore a multi-factorial set of data to explain the spatial patterns observed.

Acknowledgements

The field work, realised by A. F. Venâncio during her PhD studies, benefitted from collaboration and methodological support elaborated during the CNPq-IRD programme "Ecologie du paysage, dynamique des agroécosystèmes et complexes pathogènes: la définition du risque éco-épidémiologique dans le cas de la trypanosomose américaine, EDCTA" (2002-2004). We are thankful for the authorisation for field work in the APA by the Brazilian institutions (IBAMA, CECAV, Ministério da Saúde, Comitê de Ética em Pesquisa and Secretaria de Meio Ambiente e Recursos Hídricos do Estado da Bahia - SEMARH, Superintendência de Biodiversidade, Florestas e Unidades de Conservação - SFC).

The authors thank the inhabitants of Santa Rita for their participation and the SEAS-Guyane project for providing the SPOT-5 image.

References

Abad-Franch F, Ferraz G, Campos C, Palomeque FS, Grijalva MJ, Aguilar HM, Miles MA, 2010. Modeling disease vector occurrence when detection is imperfect: infestation of Amazonian palm trees by Triatomine bugs at three spatial scales. PLoS Negl Trop Dis 4, e620.

Barbu C, Dumonteil E, Gourbière S, 2010. Characterization of the dispersal of non-domiciliated *Triatoma dimidiata* through the selection of spatially explicit models. PLoS Negl Trop Dis 4, e777.

Bellier E, Monestiez P, Durbec J-P, Candau J-N, 2007. Identifying spatial relationships at multiple scales: principal coordinates of neighbour matrices (PCNM) and geostatistical approaches. Ecography, 385-399.

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H, 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85, 1826-1832.

Costa J, Peterson AT, Beard CB, 2002. Ecologic niche modeling and differentiation of populations of *Triatoma brasiliensis* Neiva, 1911, the most important Chagas' disease vector in northeastern Brazil (Hemiptera, Reduviidae, Triatominae). Am J Trop Med Hyg 67, 516-520.

Dray S, Legendre P, Peres-Neto PR, 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol Model 196, 483-493.

Dumonteil E, Gourbiére S, 2004. Predicting *Triatoma dimidiata* abundance and infection rate: a risk map for natural transmission of Chagas disease in the Yucatan peninsula of Mexico. Am J Trop Med Hyg 70, 514-519.

GRASS Development Team, 2010. Geographic Resources Analysis Support System (GRASS GIS) Software. Open Source Geospatial Foundation. USA. http://grass.osgeo.org (accessed on December 2010)

Khan OA, Davenhall W, Ali M, Castillo-Salgado C, Vazquez-Prokopec G, Kitron U, Soares Magalhaes RJ, Clements A, 2010. Geographical information systems and tropical medicine. Ann Trop Med Parasitol 104, 303-318.

Kitron U, Clennon JA, Cecere MC, Gürtler RE, King CH, Gonzalo M, Vazquez-Prokopec G, 2006. Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. Geospat Health 1, 49-58.

Legendre P, Eugene Gallagher E, 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271-280.

Pagès J, 2002. Analyse factorielle multiple appliquée aux variables qualitatives and aux données mixtes. Rev Stat Appl 50, 5-37.

Peterson AT, 2006. Ecologic niche modeling and spatial patterns of disease transmission. Emerg Infect Dis 12, 1822-1826.

Peterson AT, 2007. Ecological niche modelling and understanding the geography of disease transmission. Vet Res 43, 393-400.

Peterson AT, Sánchez-Cordero V, Beard CB, Ramsey JM, 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. Emerg Infect Dis 8, 662-667.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ramirez-Sierra MJ, Herrera-Aguilar M, Gourbière S, Dumonteil E, 2010. Patterns of house infestation dynamics by non-domiciliated *Triatoma dimidiata* reveal a spatial gradient of infestation in rural villages and potential insect manipulation by *Trypanosoma cruzi*. Trop Med Int Health 15, 77-86.

Romaña CA, 2004. Eco-épidémiologie. In: Lecourt, Dictionnaire de la pensée médicale, Presses Universitaire de France (PUF), 378-382.

Romain G, Roux E, Berger F, Stefani A, Gaborit P, Carinci R, Issaly J, Carme B, Dusfour I. 2011. Unraveling relationships between *Anopheles darlingi* (Diptera: Culicidae) densities, environmental factors and malaria incidences: understanding variable patterns of transmission in French Guiana (South America). Ann Trop Med Parasitol 105, 107-122.

Roux E, de Fátima Venâncio A, Girres J-F, Romaña C, 2009. High resolution remote sensing and heterogeneous data analysis for local screas. IN: XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, pp. 7577-7586.

Silveira AC, Rezende DF, 1994. Epidemiologia e controle da transmissão vetorial da doença de Chagas. Rev Soc Bras Med Trop 27, 11-22.

Slimi R, El Yacoubi S, Dumonteil E, Gourbière S, 2009. A cellular automata model for Chagas disease. Appl Math Model 33, 1072-1085.

Wagner, 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84, 1045-1057.