

# Hierarchical spatial modeling of the presence of Chagas disease insect vectors in Argentina. A comparative approach

Pablo Juan<sup>1</sup>, Carlos Díaz-Avalos<sup>2</sup>,  
Nancy R. Mejía-Domínguez<sup>3</sup> and  
Jorge Mateu<sup>1</sup>

Received: date / Accepted: date

<sup>1</sup> Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón, Spain.

<sup>2</sup> Research Institute in Applied Mathematics and Systems, Autonomus National University of Mexico, Mexico D.F.

<sup>3</sup> Red de Apoyo a la Investigación. Coordinación de la Investigación Científica, Universidad Nacional Autónoma de México, Mexico D.F.

## Abstract

We modeled the spatial distribution of the most important Chagas disease vectors in Argentina, in order to obtain a predictive mapping method for the probability of presence of the vector species. We analyzed both the binary variable of presence-absence of Chagas disease and the vector species richness in Argentina, in combination with climatic and topographical covariates associated to the region of interest. We used several statistical techniques to produce distribution maps of presence-absence for the different insect species as well as species richness, using a hierarchical Bayesian framework within the context of multivariate geostatistical modeling. Our results show that the inclusion of covariates improves the quality of the fitted models, and that there is spatial interaction between neighboring cells/pixels, so mapping methods used in the past, which assumed spatial independence, are not adequate as they provide unreliable results.

**Keywords:** Binary spatial data, Chagas vector, Covariate and hierarchical Bayesian modeling.

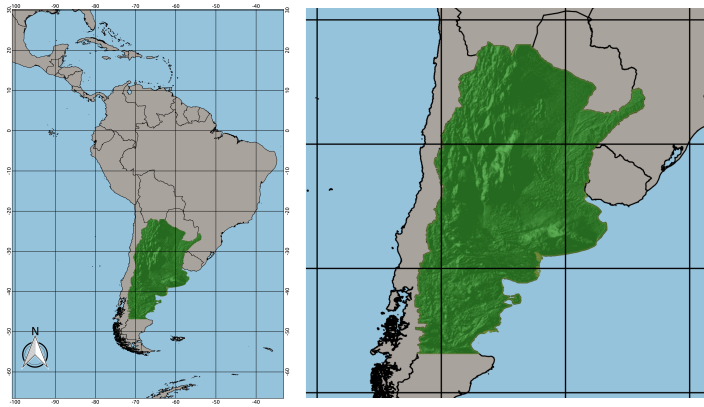
## 1 Introduction

Chagas disease is caused by infection with the protozoa *Trypanosoma cruzi*, transmitted by triatomine bugs. The disease has been considered an endemic to 22 countries in the continental Western Hemisphere (WHO, 2002, [47]) and is one of the most important

---

Corresponding author. *Email address:* juan@uji.es.

Address(es) of author(s) should be given



**Fig. 1.** Map showing the location of Argentina in South America, and a zoom of the study area, showing the topographic features of the terrain.

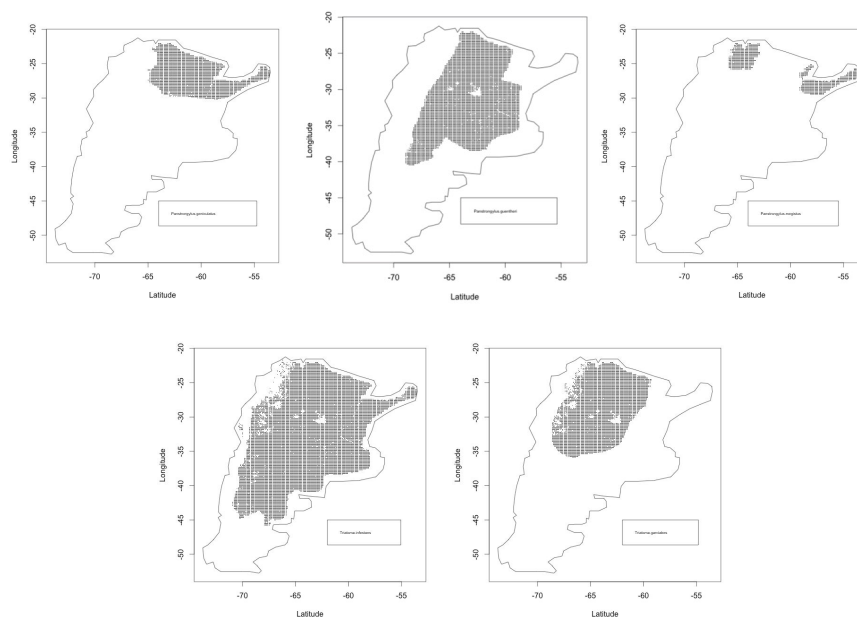
vector-borne diseases, with variable prevalence rates in tropical and subtropical areas of America. The study region of our study is Argentina, as shown in Figure 1. The estimated rate of prevalence for different countries ranges between 0.7% and 15.5%, and for Argentina, the estimated prevalence rate of Chagas disease is 8.2% (Schmunis and Yadon, 2010 [43]).

Among the 17 most common vector species related to Chagas disease, *Triatoma*, *Panstrongylus* and *Psammolestes* bugs are perhaps the most important and widespread vectors of *Trypanosoma cruzi*, the causative agent of Chagas disease. These vectors are widely distributed in Argentina and other South American countries, where they probably contribute to more than a half of the estimated 24 million cases of this disease (WHO, 2002, [47]). However, there are few formal studies regarding the geographic distribution of each species and its relationship to climatic and geographic factors related to their presence at a given location. Most of the previous studies in this regard have used ecological niche models or simple mapping and direct density computations (Cruz-Reyes and Pickering-López, 2006, [13]) ; (Gurgel-Goncalves et al, 2012, [20]). None of these studies have considered the presence of spatial association between neighboring areas in the species distribution nor the possibility of local variability in the presence of the vector insect species.

In this paper we analyzed the spatial distribution of five vector species considered the most important in the spreading of the pathogen. Our goal is to enable predictive mapping of the spatial distribution of the probability of presence of the insect vector species. The probability maps constructed may be used by experts in Chagas disease transmission to link the presence of the vectors to the probability of presence of the disease (Adin et al., 2016 [1] and Chia et al., 2014 [9]). Such information is highly useful in planning health care and disease control activities. We analyzed the binary variable of presence-absence of Chagas disease vectors in Argentina, using a data base obtained from a long term field survey on a grid covering the northern part of the country, including climatic and topographical covariates. Our analyses comprised the use of several statistical techniques designed to produce distribution maps of probability of presence of a given vector at specific locations within a grid covering the Argentinean territory (Zeledon and Rabinovich, 1981, [48]). We modeled the presence-absence data and species richness using a hierarchical Bayesian framework within the context of autologistic regression models (Besag, 1974 [3] or Ugarte et al., 2005 [46]), ranging from simple logistic regression models, as a particular case, to autologistic regression models.

We also explored models to test the possibility of spatial interaction between different vector species.

The plan of the paper is the following. In the Section 1.1, we presents the data set with the vectors of Chagas disease and its preparation to further modeling. In Section 2 we present the statistical methodology used to fit autologistic regression models based on a Bayesian framework. We develop several competing models and compare them with the results obtained with the SPDE-INLA method. Section 3 discusses the results, and a discussion is given in Section 4. The paper ends with a final section with conclusions. An Appendix is added with tables and figures for completeness.



**Fig. 2.** Presence data per species. *First row: Panstrongylus geniculatus, Panstrongylus guentheri and Panstrongylus megistus. Second row: Triatoma infestans and Triatoma garciabesi.*

The study area comprises the central and northern part of Argentina, a country located in South America (Figure 1). Most of Argentina's territory is located in low altitude lands, except for the west and northwest regions which are areas of higher altitude. It is in these areas where the Andes mountain range is located, with elevations above 4000 meters. Going in the east-west direction one finds an altitudinal gradient that has a strong influence on the climatic variables. In the south-north direction, changes in latitude define a pattern of increasing values of Normalized Difference Vegetation Index (NDVI), explained mostly by the distribution of moisture and temperature. In the northeastern part of the country we can find the most tropical area with the highest rainfall and largest LST values. Near the center of the continental region, the average annual rainfall decreases, but is "compensated" by the relative humidity coming from the Atlantic Ocean. This climatological and physiographic mosaic makes Argentina a zone with conditions adequate for the spread of the insects that may transmit Chagas disease to humans and to domestic animals.

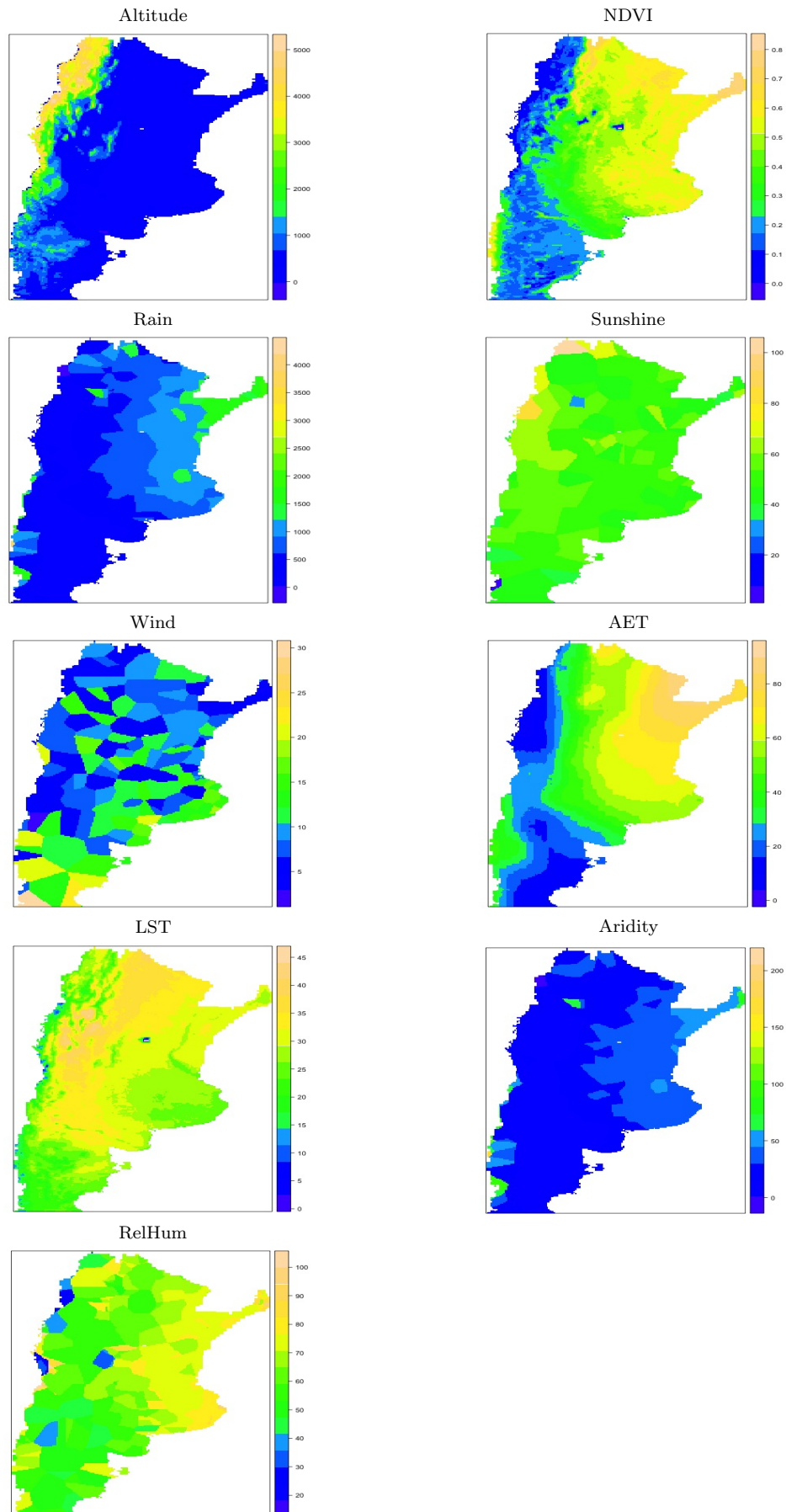


Fig. 3. Covariates used in the modeling process.

## 1.1 Data

The data base we analyzed consisted of presence and absence data obtained from a major field survey on a grid of  $0.1 \times 0.1$  degrees (about  $10 \times 10$  Km pixels). For a given pixel and a given species, the presence was recorded if the species was reported as present at any location within the pixel, regardless of the number of individuals. The data base was kindly provided by J. Ravinovich, from Centro de Estudios Parasitologicos y de Vectores (CONICET, CCT- La Plata, UNLP), Universidad Nacional de La Plata, Argentina. This data set corresponds to periodic (36 months) census made between 1980 to 2001. We used the mean values for presence-absence data, and also for covariates. Using Geographic Information Systems (GIS), digital maps for altitude, NDVI, Rainfall, sunshine, wind, actual evapotranspiration (AET), land surface temperature (LST), aridity and humidity for the entire Argentinean territory were obtained at the same spatial resolution as the presence data, to be included as covariates in the modeling process. All the covariate raster data were converted to the coarser resolution of 10X10Km in which the presence data were obtained. This scale is adequate for our objective, since we want to find the factors affecting distribution of the insect vector species at a country scale. Anyway, the reader should be aware that changes of scale change many of the inferences in an analysis (Qi and Wu, 1996 [35]).

The altitude for the study area is between 80-4980 m, with an average of 709 m. The mean value for NDVI for the region is 0.38 (max = 0.79), mean value for precipitation (Rain) is 621 mm with a great variation: minimum value is 12 mm and maximum is 4199 mm. Also, Land Surface Temperature shows a  $44.3^{\circ}\text{C}$  as maximum value but  $2.4^{\circ}\text{C}$  as minimum, the mean value is  $29.5^{\circ}\text{C}$ . Relative Humidity average is 62.14%, but the highest value is 100 %. For the Aridity Index, the maximum value is 205, however, the mean value is 20.3 with a standard deviation equal to 14.9. Also, for sunshine, the mean is  $50.5 \pm 8.8$  standard deviation. Finally, the mean value for AET is  $44.6 \pm 21.5$  standard deviation.

The raw presence data for the 5 insect vector species of Chagas disease are shown in Fig. 2. It can be seen that the three species of the genus *Panstrongylus* were detected only in a relatively small area in the northern part of the country, in warm and humid environments. On the other hand, *Triatoma infestans* and *Triatoma garciabesi* were recorded in a larger area, which includes a wider range of temperatures and humidity. Note that none of the five species was recorded in the areas near the Andes , where dry and cold conditions prevail.

Chagas disease vectors are widely distributed in Argentina and other South American countries, where they probably contribute to more than a half of the estimated 24 million cases of this disease. One of these field observational studies records data on presence-absence of the above mentioned disease vectors for a variety of species over a fine grid covering part of Argentina. In addition, climatic and topographical covariates were also measured to assess their effect on the spatial distribution of the insect vector species considered in our study and to test their significance in spatial distribution predictive maps. The covariate maps for the study area are shown in Figure 3. Some of the maps show a mosaic shape because they could not be obtained at a finer resolution; yet, they provide limits on their spatial variability and their contribution to explain the spatial variability of the species distribution.

## 2 Statistical modeling

### 2.1 Autologistic regression models

The use of the raw observations to produce spatial distribution maps may be misleading because of the high variability inherent to samples of spatial variables (Besag, 1991, [4]). To make inferences about a random phenomenon, statistical models have proven to be a sensible approach. Because the data available for this study were in raster format, statistical inferences about the quantities of interest may be done using Markov Random Field models. A raster image of  $n = I$  rows and  $J$  columns is equivalent to a lattice with  $n = I \times J$  nodes, where each node represents a pixel in the image. Since we may label each node with an integer number, lattices are not restricted to regular shapes and can be applied to areas irregularly shaped.

For data defined in a lattice with  $n$  nodes or vertices, as in our case, functionals of the spatial distribution of the variable of interest such as the mean, median or maximum for instance, given the values observed at the rest of the image,  $P(Y_i|Y_{-i}, \theta)$ , as well as the parameters of the distribution of  $Y$  are the target for estimation and statistical inference.  $Y_{-i}$  refers to the values of  $Y$  in all the lattice except the  $i$ -th node and  $\theta$  is a vector of model parameters for  $P(\cdot|\cdot, \theta)$ . Markov Random Field (MRF) theory allows stochastic modeling of many spatial phenomena based on the conditional distribution of the variable of interest over a set of spatial locations in a lattice, and were described in detail by Besag (1974) [3]. Besag (1974) ([3]) described in detail the class of auto-models. MRFs allow to make inferences for a wide range of applications in spatial modeling, including discrete and continuous spatial variables. In particular, for the class of auto-models, the spatial variable of interest  $Y$  may follow any distribution in the exponential family and shows some degree of association with the values of  $Y$  at a set of neighbouring locations. The exponential family of distributions includes the binomial, the Poisson and the Gaussian among others. This gives the auto-models a great flexibility to model many spatial phenomena.

Besag (1974) proved that under some mild conditions,

$$P(Y_i|Y_{-i}, \theta) = P(Y_i|Y_{\sim i}, \theta)$$

where  $Y_{\sim i}$  denotes the values of  $Y$  at those locations (pixels) that are neighbours of location  $i$ . A detailed description of the properties of MRF is beyond the scope of this paper. For a description on the derivation of auto-models, the interested reader is referred to the seminal paper by Besag (1974, [3]) and to the book by Cressie (1993) [12].

When modeling the spatial distribution of dichotomic random variables the aim is to produce probability maps. In our study, the goal is to produce maps showing the probability of presence of the vector species at any given pixel within the study area, which we will denote by  $\mathcal{D}$ . Note that we assume that the species are absent when the data base recorded a zero, although it is possible that a zero has been recorded in places where the species are not observed but are still present in the area covered by the corresponding pixel in the image. Let  $Y_i$  be a Bernoulli random variable taking the values  $Y_i = 1$  if the insect vector species is observed at pixel  $i$  and  $Y_i = 0$  otherwise. We define  $p_i = P[Y_i = 1]$  which may be interpreted as the probability that the vector species has pixel  $i$  within its spatial distribution zone. Our interest here is to estimate the  $p_i$  and to test their possible association with a set of covariates  $Z_1, \dots, Z_K$ .

Given the binary nature of the data  $y_i$ , a natural choice is to fit logistic regression models of the form

$$\xi_i = \text{logit}(p_i) = \log \left\{ \frac{p_i}{1 - p_i} \right\} = z_i^T \beta$$

to the presence-absence data of each of the five species considered as the most important.  $\beta$  is a vector whose entries are the coefficients related to the spatial covariates

**z.** Logistic regression assumes that the data are independent, but such assumption might not be adequate if the data are not spatially independent. We thus alternatively fit the model

$$\xi_i = \mathbf{z}_i^T \beta + \psi_i \quad (1)$$

where  $\psi_i = \sum_{i \sim j} \psi_j$  is a term that incorporates the spatial interaction between neighboring areas and may be seen as a surrogate for unobserved covariates that are correlated in space (Besag *et al.*, 1995 [5]). This approach corresponds to generalized linear mixed models (GLMM), a class of models useful in problems that involve the mapping of risks (Clayton and Kaldor, 1987 [10]). From (1) we obtain

$$p_i = \frac{\exp \{y_i x_i' \beta + \psi_i\}}{1 + \exp \{y_i x_i' \beta + \psi_i\}}$$

In a similar fashion, to model the spatial distribution of species richness we fitted the model

$$E[X_i] = z_i^T \beta + U_i \quad (2)$$

In models (1) and (2) the spatial terms  $\psi_i$  and  $U_i$  will be assumed to be Gaussian Markov Random Fields (GMRFs). Thus, model (1) corresponds to a logistic normal model (Díaz-Avalos *et al.*, (2001) [15]) whilst model (2) corresponds to an ordinary linear regression model with spatial dependent errors.

Our approach to model bases on the Bayesian context, that is, if  $\theta$  is a vector of  $k$  components containing all the parameters in the model, statistical inferences about  $\theta$  are based on the posterior distribution

$$f(\theta|y) = \frac{L(\theta; y)\pi(\theta)}{\int_{\Theta} L(\theta; y)\pi(\theta)d\theta} \propto L(\theta; y)\pi(\theta)$$

Our model formulation is as follows. We assumed a flat, noninformative prior distribution (Box and Tiao, 1973 [7]) for the nonspatial parameters in our model, which allows them to be assigned any arbitrary initial values. Following Besag *et al.* (1991 [4]), we assumed that  $\psi$  is a markov random field with a Gaussian pairwise difference prior distribution, with precision  $\lambda$ , this is,

$$p_i(\psi) \propto \lambda^{0.5N} |W|^{0.5} \exp\{-0.5\lambda\psi'W\psi\} \quad (3)$$

where  $\psi = (\psi_1, \dots, \psi_N)$  is the vector of spatial components,  $W$  is a matrix with  $W_{ii} = \nu_i$ ,  $W_{ij} = 1$  if pixels  $i$  and  $j$  are neighbors and  $W_{ij} = 0$  otherwise. We are weighting each direction equally, because the geographic scale we are working with is not detailed enough to detect possible anisotropies. The prior density previously described belongs to the class of nonstationary Gaussian intrinsic autoregressions, and may be considered as the stochastic equivalent of linear interpolation (Besag *et al.*, 1991 [4]). The Gaussian conditional autoregressive processes have been used extensively in spatial statistics and Bayesian image analysis, both for regular and irregular arrays. Gaussian conditional autoregressive model allows some smoothing in the predictive maps and maintains the autoregressive structure of the CAR model. Also, the use of the GCAR model prior for the spatial effect in our model makes the full conditional distributions computationally simple, something desirable when using MCMC methods.

Because both the columns and rows of  $W$  add to zero, this prior is improper. However, the full conditional densities necessary to make statistical inferences on the  $\psi_i$ ,  $i = 1, \dots, N$  are well defined (Besag, *et al.*, 1995 [5]). For the precision  $\lambda$  we assumed a  $G(1, 1)$  prior density, which allows initially low values for  $\lambda$  and, therefore, high variability in  $\psi$ . The model is completely specified by further assuming independence between the

components of  $\theta = (\beta, \psi, \lambda)$  and by assuming that the observations  $y_i$  are conditionally independent. The posterior density of the parameters is proportional to

$$\left( \prod_{i=1}^N \frac{\exp\{y_i x_i' \beta + \psi_i\}}{1 + \exp\{y_i x_i' \beta + \psi_i\}} \right) \times \lambda^{0.5N} |W|^{0.5} \exp\{-0.5\lambda \psi' W \psi\} \times \lambda^{a-1} \exp\{-b\lambda\}$$

and the full conditional distributions are given by

$$\begin{aligned} \pi(\beta|\cdot) &\propto \left( \prod_{i=1}^N \frac{\exp\{y_i x_i' \beta + \psi_i\}}{1 + \exp\{y_i x_i' \beta + \psi_i\}} \right) \\ \pi(\psi|\cdot) &\propto \left( \prod_{i=1}^N \frac{\exp\{y_i x_i' \beta + \psi_i\}}{1 + \exp\{y_i x_i' \beta + \psi_i\}} \right) \times \exp\{-0.5\lambda \psi' W \psi\} \\ \pi(\lambda|\cdot) &\sim \Gamma(a + 0.5N, b + \sum_{i=1}^N \nu_i (\psi_i - \bar{\psi})^2) \end{aligned}$$

where the notation  $(u|\cdot)$  means the conditional distribution of one set of parameters given the rest of the components in the model.

We ran the MCMC algorithm described previously using random updates of the full conditional distributions. Although convergence of the MCMC depends on the proportion of positive to null observations, we used a minimum of 3000 simulations as "burning up" and left the algorithm run until 10 000 iterations were achieved. At the end of each iteration, a "posterior" probability of presence was also computed for each pixel. We used the results of the last 7000 iterations to compute posterior means, medians and quantiles of posterior probability of presence, as well as for the rest of the model parameters.

## 2.2 INLA and SPDE approach

MCMC has the disadvantage of requiring intensive computations that in many cases make the algorithm slow and time consuming. Although advances in computer science and electronics have made this a rare issue, it is sometimes desirable to get the results in short time. One methodology based on INLA and the stochastic partial differential equation (SPDE) approach provides quicker results. The SPDE approach allows a Gaussian field with Matérn type covariance function

$$M(h | \nu, \kappa) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|h\|)^\nu K_\nu(\kappa \|h\|) \quad (4)$$

to be approximated as a discretely indexed spatial random field. Further use of the integrated Laplace transform produces significant computational advantages (Lindgren et al. (2011) [30]). In (4)  $K_\nu$  is a modified Bessel function of the second kind and  $\kappa > 0$  is a spatial scale parameter whose inverse,  $1/\kappa$ , is sometimes referred to as a correlation length. The smoothness parameter  $\nu > 0$  defines the Hausdorff dimension and the differentiability of the sample paths (Gneiting et al. (2010) [19]). Specifically, we tried  $\nu=1,2,3$  (Plummer (2008) [34]). Using the expression defined in the previous formulae, when  $\nu+d/2$  is an integer, a computationally efficient piecewise linear representation can be constructed by using a different representation of the Matérn field (see equation 5), namely as the stationary solution to the stochastic partial differential equation (SPDE) (Simpson D. et al., (2001) [42]).

The idea behind INLA and SPDE is to construct a finite representation of a Matérn random field by using a linear combination of basis functions defined in a triangulation



of a given domain  $D$ . This representation gives rise to the SPDE, which is a link between the GF and the GMRF. Such link allows the spatial or the spatio-temporal covariance function and the dense covariance matrix of a GF to be replaced by a neighborhood structure and a sparse precision matrix, respectively, both of which are typical elements that define a GMRF, which in turn yields substantial computational advantages (Lindgren et al. (2011) [30]).

In particular the SPDE approach consists in defining the continuously indexed Matérn GF,  $U(s)$  as a discrete indexed GMRF by means of the representation of a basis function defined on a triangulation of the domain  $D$

$$U(s) = \sum_{l=1}^n \varphi_l(s) \omega_l \quad (5)$$

where  $n$  is the total number of vertices in the triangulation,  $\{\varphi_l(s)\}$  is the set of basis functions, and  $\{\omega_l\}$  are zero-mean Gaussian distributed weights. The basis functions are not random, but are instead chosen to be piecewise linear on each triangle

$$\varphi_l(s) = \begin{cases} 1 & \text{at vertex } l \\ 0 & \text{elsewhere} \end{cases} \quad (6)$$

The key is to obtain the weights  $\{\omega_l\}$ , which report on the value of the spatial field at each vertex of the triangle. The values inside the triangle will be determined by linear interpolation (Simpson et al. (2011) [42] and developed by R-INLA project (2012) [37]).

Thus, expression (5) defines an explicit link between the Gaussian field  $U(s)$  and the Gaussian Markov Random Field, and it is defined by the Gaussian weights  $\{\omega_l\}$  that can be given by a Markovian structure. The Matérn spatial covariance function controls the spatial correlation at distance  $\|h\| = \|u_i - u_j\|$ , where  $u_i$  and  $u_j$  are two arbitrary locations within a spatial domain or study area, separated by a distance  $\|h\|$ .

When  $\nu + d/2$  is an integer in (4) a computationally efficient piecewise linear representation can be constructed by using a different representation of the Matérn random field  $u(i)$ , namely as the stationary solution to the SPDE (Simpson *et al.* [42])

$$(\kappa^2 - \Delta)^{\alpha/2} U(i) = W(i) \quad (7)$$

where  $\alpha = \nu + d/2$  is an integer,  $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial i^2}$  is the Laplacian operator, and  $W(i)$  is spatial white noise.

In a statistical analysis, to estimate a general model it is useful to model the mean for the  $i$ -th unit using an additive linear predictor, defined on a suitable scale

$$\eta_i = \beta_0 + \sum_{m=1}^M \beta_m z_{mi} + \sum_{l=1}^L f_l(v_{li}) \quad (8)$$

Here  $\beta_0$  is a scalar which represents the intercept,  $\beta = (\beta_1, \dots, \beta_M)$  are the coefficients which quantify the effect of some covariates  $z = (z_1, \dots, z_M)$  on the expected value of the presence-absence random variable, and  $f = \{f_1(\cdot), \dots, f_L(\cdot)\}$  is a collection of functions defined in terms of a set of random effects  $v = (v_1, \dots, v_L)$ . In our particular case,  $v = (v_1, \dots, v_L)$  are the conditional means associated to the GMRF. Changing the form of the functions  $f_l(\cdot)$  allows to fit different kind of models, from standard and hierarchical Bayesian regression, to more complex spatial models (Rue et al., 2009 [40]). Given the specification in (8), the vector of parameters is represented by  $\theta = \{\beta_0, \beta, f\}$ . Unlike the MCMC algorithm, which needs to evaluate the full conditional distributions at each iteration of the algorithm, the INLA approach exploits the model assumptions to produce a numerical approximation to the posteriors distributions of interest, based on

the Laplace approximation over a tessellation constructed in a grid (Tierney and Kadane (1986) [45]).

The use of INLA and the SPDE algorithms yields significant savings in computational time and allows the user to work with relatively complex models in an efficient way. In our application, all the analyses are carried out using the R freeware statistical package (version 3.1) (R Development Core Team (2011) [36]) and the R-INLA package (R-INLA project (2012) [37]). To test the robustness of our methodological choice we used several non-conjugate priors for the parameters, in particular Gaussian and flat priors for the model parameters. We found no notorious changes in the empirical posterior distribution for the precision hyper-parameters.

We have preferred to use the corresponding posterior distributions assuming a Binomial likelihood with non-conjugate priors for the model parameters. A detailed technical description of the INLA and SPDE approach is beyond the scope of this paper. The interested reader is referred to review Lindgren et al. (2011) [30] and Rue, Martino, Chopin 2009 [40] or Illian et al. 2013 [26]. As we mentioned previously, the autologistic models were fitted using Markov Chain Monte Carlo (MCMC) whilst for the INLA method we used the Stochastic Partial Differential Equation (SPDE) approach, which is the building block for the Integrated Nested Laplace Approximation (INLA) library for the R software for the autologistic regression models (Lindgren and Rue (2013) [29], Lindgren et al. (2011) [30], Rue and Martino (2006) [38], Rue et al. (2007) [39], Rue et al. (2009) [40]).

### 3 Results

The results for the different fitted models in this study are shown in Table 1, where we show the marginal likelihood and in parenthesis the CPO score or cross validated score, defined as the mean of the negative of the conditional probability  $\pi(y_i|y_{-i})$ . Scoring rules are negatively oriented, which means that, the smaller the score, the better the predictive power of the model (Schrodle et al. 2010). An attractive feature of this measure is that it can be applied to parametric and non-parametric settings and does not require models to be nested, nor to be related in any way (Gneiting and Raftery, 2007). According to this, except for the case of *P. megistus*, the inclusion of covariate information improves the predictive performance of the models, meaning that part of the spatial variation in the probability of presence of the insect vectors considered in this study is explained by the spatial variation of climatic and biological factors.

Tables 2 to 4 show a summary of the posterior statistics for the parameters of the fitted models using the MCMC approach, whilst Tables 5 to 7 show the corresponding posterior estimates for the fitted models using the INLA approach. The tables only show those terms that were statistically significant. The coefficient estimates in the tables show that the logit for the presence of *P. geniculatus* is positively associated to rain, AET and LST, terms that were significant both with MCM and INLA methods (Tables 2 and 5). Climatic factors were significant only for the MCMC method.

According to these coefficients, the probability of presence for *P. geniculatus* is inhibited by increments in altitude, sunshine and humidity. For *P. guentheri*, the coefficients for altitude, NDVI, AET and LST were significant for the full fitted model by the two methods. In both cases, the logit of its probability of presence at a given location in Argentina is positively associated to changes in NDVI, AET and LST. By the INLA method the logit of probability of presence is negatively associated to increments in rain and humidity. The third species of this genus, *P. megistus* has a logit of probability of presence associated positively to changes in NDVI and humidity (Tables 3 and 6).

The posterior probability of presence maps for a logistic model of *Triatoma* genus shows a positive association with humidity, in contrast with the genus *Panstrongylus*

whose distribution includes zones with lower humidity even with highest aridity and high AET (Figures 5 to 8). Particularly, *P. geniculatus* prefers the lowlands, specifically those zones with tropical savannas in the north of Argentina. *P. guentheri* is essentially distributed in dry areas, i.e., with little rainfall, low humidity and high evapotranspiration. *P. megistus* shows a restricted distribution that corresponds with Atlantic forest present in this country, in this case indicated for higher values of NDVI. For *Triatoma* species, the posterior probability of presence maps predicted largest areas of distribution. *T. infestans* is the most widely distributed species. This species is absent in highlands but in high humidity and temperature. Also, *T. garciabesi* is present in zones with high humidity but with high altitude, and wooded sites (Fig. 5 and 6). The posterior probability of presence for *Triatoma* richness shows that the high values corresponded to the zones with high evapotranspiration but medium relative humidity (Fig. 7 and 8). Similarly, total species richness shows the same pattern, because it shows an extended distribution.

The posterior probability of presence maps as well as the standard deviation maps for the five species considered in his study are shown in Figs. 5 to 8, both obtained with INLA and MCMC methods. Although both methods agree in the shape and the extent of the spatial distribution of the five insect species under consideration, the maps show differences, particularly regarding the smoothness of the fitted random field. Fig. 9 shows the posterior mean and its standard deviation for the species richness using the five species (upper panel) and for the two triatomus species (lower panel) for both cases with and without covariates. The posterior mean estimates for the fitted models with and without covariates show the same general pattern in the study area, with high values in the central north part of Argentina, this is, both models capture the large scale variability of the spatial distribution of species richness. However, the model with covariates is able to better capture the local spatial variability of species richness, which is expected to be associated to local conditions of topography and other environmental variables, something that may be very important for health authorities at the municipalities.

#### 4 Discussion

The Bayesian approach has produced distribution maps for the insect vector species considered in this paper producing results that are consistent with what the experts in epidemiology have reported for the Chagas disease using non statistical methods. The inclusion of spatial covariates in the models not only improves the quality of the resulting maps, but also is an aid to screen the factors related to the spatial distribution of the vectors of Chagas disease. Knowledge of such factors may be used to define critical or potentially critical areas for outbreaks of insect bites and thus for an increase in Chagas disease prevalence. Thus, the statistical methods used here can be used as a tool for planning and to draw potential scenarios for the distribution of the vector insects under the global warming and climate change. The coefficient estimates associated to each covariate provide evidence about which factors are related to the spatial distribution of the vectors of Chagas disease. Clearly, one should not expect the presence of all the species to be associated with the same strength to the physical and climatic factors, as their spatial distribution is the result of a long process of competition and adaptation to different environments (Galvao and Justi, 2015 [17]).

The posterior probability maps show the almost ubiquity of the two *Triatomus* species. This is evident in the individual posterior probability of presence and in the posterior mean maps for the species of those two *Triatomus* species. As we mentioned previously, the distribution range of these species is controlled mainly by temperature and temperature-related covariates such as NDVI, LST, altitude above sea level and evapotranspiration. These results suggest that these two species have been very successful

**Table 1.** Marginal likelihood and CPO for INLA models for individual species and species richness

Species	Model	
	without covariates	with covariates
<i>P. geniculatus</i>	-478.56 (2.72)	-14238.87 (1.69)
<i>P. guentheri</i>	-1370.47 (6.54)	-1381.34 (4.82)
<i>P. megistus</i>	-420.53 (2.96)	-13889.78 (3.08)
<i>T. infestans</i>	-3163.92 (0.27)	-3102.18 (0.24)
<i>T. garciabesi</i>	-1581.68 (7.47)	-1528.10 (6.71)
Species richness	-36346.57 (1.438)	-36316.80 (1.434)
<i>Triatoma</i> richness	-38134.80 (1.520)	-38088.08 (1.516)

to survive under a wide spectrum of ecological conditions, and this explains why they are considered the most important vectors for Chagas disease at least in South America (Gurtler and Yadon, 2015 [21]; Schofield et al., 2006 [41]). Further, *T. infestans* has been described as a species highly adapted to peridomestic environments (Galvao and Justi, 2015 [17]) whilst, on the other hand, *T. garciabesi*, is seldom recorded in those environments and apparently does not colonize households either (Canale et al., 2000 [8]). *T. garciabesi* has been recorded mainly in areas with low rainfall and temperatures in the north and central parts of Argentina (Gurgel-Goncalves et al., 2012 [20]) but often associated to tree covered areas, as this species is found mainly under the tree bark (Galvao and Justi, 2015 [17]).

On the other hand, models predict a distribution of genus *Panstrongylus* in zones with lower humidity even with highest aridity and high AET. Particularly, *P. guentheri* is essentially distributed in zones occupied by "Chaco" region and savannas. This species has been recorded associated to peridomestic environments, living in between the wood used in domestic fireplaces (Galvao and Justi, 2015 [17]). The *P. geniculatus*' distribution matches with zones with tropical savannas and "cerrado" in the north of Argentina. This distribution pattern in the north of Argentina is the continuation of the distribution pattern observed in Brazil and Paraguay (Galvao and Justi, 2015 [17]), which in the boarder with Argentina share a subtropical climate. *P. megistus* shows a restricted distribution to the spatial distribution of the Atlantic forest or "cerrado" as reported by Gurgel-Goncalves et al. (2012 ([20])). For *P. megistus* both models pinpoint NDVI as the covariate with the stronger association to the posterior probability of presence at any given location, with higher NDVI values associated to higher probability of presence. The last two species have been reported as ongoing a process of adaptation to household environments (Galvao and Justi, 2015 [17]) and therefore it is likely that future studies do not show this strong association between the probability of presence and NDVI.

The MCMC method suggests a positive association of the logit of presence for this species to sunshine and AET, and a negative association to LST and humidity. *P. geniculatus* and *P. guentheri* seem to prefer the same kind of habitat, in which the vegetation coverage is high, with high evapotranspiration due in part to high temperatures in low altitude zones, whilst *P. megistus* seems to prefer the same kind of habitat, but in areas where temperatures are lower than those preferred by the two former species. Regarding the species of the genus *Triatoma*, the logit of presence for *T. infestans* in the study area is negatively associated to altitude, and positively associated to sunshine, AET, LST and humidity, whilst the logit of presence for *T. garciabesi* is positively associated to altitude, NDVI, LST and humidity (Tables 4 and 7).

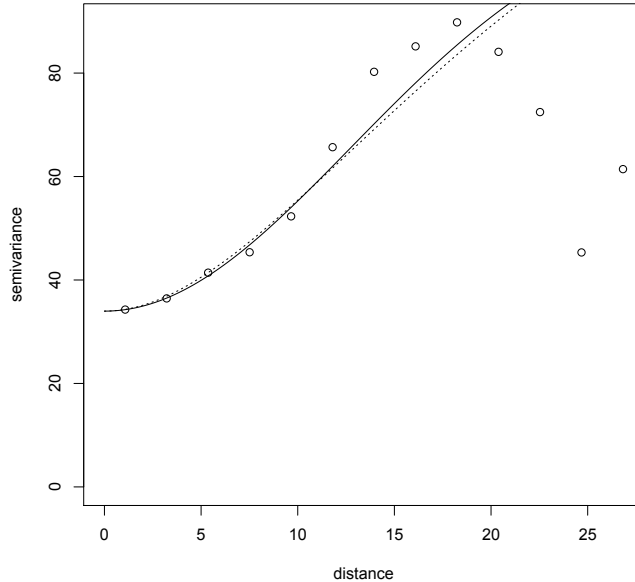
This species thus prefers warm and humid habitats, but unlike *T. infestans* and *T. garciabesi*, avoids habitats with high sunshine and high evapotranspiration. The posterior probability of presence maps obtained with autologistic models fitted using the INLA-

SPDE methodology show "abrupt" boundaries, unlike the maps obtained by fitting the autologistic models with MCMC method. This is due in part to the coarser partition made by the INLA-SPDE method and to the lack of the smoothing effect that the use of the Gaussian spatial term included in the model induces in the posterior probability maps when MCMC is used for model fitting. The standard deviation maps for both models show a dotted pattern which is a consequence of the way the INLA-SPDE package is implemented in R. Note however that high values of the standard deviation correspond to areas of high diversity as a result of the fitted Poisson model, in which the variance is proportional to the mean. Thus, areas of high species richness are predicted with higher uncertainty.

Despite the complex interactions among the abiotic factors considered in our analysis, the interrelationship between those factors and the insect vector species for Chagas disease have led to a spatial distribution pattern where only a couple of species have a wide distribution range in Argentina. The models we have fitted provide valuable information both, to assess which covariates and how they relate to the probability of presence of the five insect vector species considered in this study, and also provide a spatial map for the distribution of such probabilities. The inclusion of the spatial term in the autologistic model corrects the underestimation of the confidence intervals for the coefficients associated to the covariates, a common error when the spatial locations are assumed spatially independent. Fitting models with covariates to the spatial distribution of the probability of presence will allow to assess the changes associated to those spatial distribution patterns associated to global warming and climate change. Medone et al. (2015) [31] for instance, mention that the spatial distribution range for *T. infestans* will shrink to more temperated places as the global temperature increases. This will allow species such as *T. garciabesi* increase its distribution range according to our results. Ignoring the climate change scenario, Schofield et al. (2006) [41] predicted that *T. infestans* will spread to almost all the Argentinean territory, spreading the incidence of Chagas disease to new areas in the southern part of America. Gurtler and Yadon (2015, [21]) highlight the importance of controlling the presence of insect vectors in domestic environments, but also highlight the importance of ecological factors on the dispersion of the disease. Thus, policies for vector control should be planned to face this future scenario.

Mathematical models can then provide highly valuable tools to describe, understand and predict the presence and/or abundance of vectors, as well as to estimate the potential effect of various vector control strategies (Nouvellet et al., 2015 [32]). There are several approaches to tackle the task of producing spatial distribution maps for biological species, some of them widely used in the ecologists community. Methods such as GARP, MaxEnt and Bioclim have become the standard tool in government agencies for natural resources management, scientists working on biodiversity, ecological modeling and epidemiology (Gorla, 2002 [18], Gurgel-Goncalves et al., 2012 [20], Medone et al., 2015 [31] and Peterson et al., 2002 [33]). All these methods are based on multivariate techniques such as principal component analysis, combined with some classification rules. GARP, for instance, uses logistic classification rules and genetic algorithms to produce distribution maps. Their performance was compared with the Bayesian approach using MCMC by Díaz-Avalos (2007) [16] under a combination of spatial sampling schemes and sample sizes. He found that in general, the Bayesian approach performs better, and concluded that the Bayesian methods have a superior ability to capture small scale variation. The INLA approach used here is based also on a Bayesian setting and in principle provided similar results to those obtained with the MCMC approach, with the benefit of an easier implementation. Our resulting distribution maps differ of those reported by de la Vega *et al.* (2015) [14], but most of the differences are explained by the different spatial resolution of their study.

Formal comparisons between the INLA approach and the MCMC approach are needed in the context of binary random fields with covariates.



**Fig. 4.** Empirical variograms for spatial term  $\psi$  and theoretical models, Matern (dotted line) and Gaussian (solid line).

INLA and MCMC methods agreed in the shape and the extent of the spatial distribution in general for each species, but we observed slight differences particularly regarding the smoothness of the fitted random field. The main reason for this comes from the prior distribution considered under each method. The INLA approach assumes a continuous random field a priori for the spatial component of the model, whilst the MCMC approach considers a Markov random field. Thus, the posterior spatial distribution maps for each method differ in the degree of smoothness in the prior for the spatial component of the fitted model. It is not clear which prior produces better results, but in the context of probability of presence maps one should expect small scale variation, something that seems to be captured better by the MCMC method. Also, the way the INLA method works, with the SPDE approach and the way the tessellations for the estimation process are constructed may explain the smoothness similarity that can be observed between the probability maps with INLA and the spatial dispersion of the presence records. Previous studies comparing INLA and MCMC have highlighted these dissimilarities when the same model is fitted by those methods to the same data set (Taylor and Diggle (2014) [44]). For instance, with sparse binary data MCMC needs a long burning up time and many iterations to converge to the posterior distribution. Further, it is very sensitive to the presence of collinearity in the covariates, particularly when the proportion of presence-absence in the data is very low.

The INLA method on the other hand, does not need such burn-in time but is also sensitive to the spatial association (SPDE) created for numerical solution of the Laplace approximation. Fig. 4 shows two different theoretical models fitted to the spatial term  $\psi$

for comparison. The MCMC method does not assume any particular covariance structure for the spatial term, whilst the latent random field structure assumed by INLA has to have a Matérn type covariance. For the logistic models we assumed a Gaussian Markov random field structure for  $\psi$ , which explains the similarity of the fitted variogram models to the empirical variogram of the spatial term  $\psi$  shown in Figure 4.

Although the INLA and the MCMC work with the Bayesian framework, there are key differences that result in differences in the results with each method. A conspicuous difference between INLA and MCMC relates to the way both methods estimate the spatial association. On one hand, INLA uses the SPDE to obtain a smooth approximation between neighboring tessellation cells. On the other hand, MCMC uses the neighbor structure imposed by the modeler, resulting in noisier maps, depending of whether a first or higher order neighbor structure is used ([6]). This explains in part the discrepancies observed in the parameter estimated with both methods. A formal comparison of these methods and the non-statistical methods currently used in most instances is needed in the context of binary random fields with covariates. Such comparison should be based on the use of virtual species, as this approach allows the evaluation of the error rates and makes better comparisons between the different methods for distribution mapping.

## 5 Concluding remarks

In the modeling of spatial distribution of Chagas disease, part of the spatial variation could be modeled through the inclusion of spatially varying covariates. The improvements in the map constructed can be evaluated using any adequate statistic such as the posterior deviance or the Bayesian Information Criterion (BIC), providing potential users of the methodology shown in this paper a way to screen which covariates can be useful in their particular studies.

The Bayesian approach has produced distribution maps for the insect vector species considered in this paper that are consistent with what the experts in epidemiology have reported for the Chagas disease using non-statistical methods. The inclusion of spatial covariates in the models not only improves the quality of the resulting maps, but is an aid to screen the factors related to the spatial distribution of the vectors of Chagas disease. Knowledge of such factors may be used to define critical or potentially critical areas for outbreaks of insect bites and thus for an increase in Chagas disease prevalence. The increasing temperatures in the planet suggest that the distribution range of the *Triatomus* species will turn wider, allowing them to colonize some of the upper parts in the Andes range. Thus, the statistical methods used here can be used as a tool for planning and drawing potential scenarios for the distribution of the vector insects under a global warming and climate change.

MCMC and INLA both use a Bayesian approach, the results presented here show some differences, mainly on the estimated parameter of the logistic term and on the smoothness of the resulting maps. Such differences apparently come from the way both approaches model the spatial dependence in the data.

## Acknowledgments

We thank J. E. Rabinovich from Centro de Estudios Parasitológicos y de Vectores of Buenos Aires, Argentina for drawing our attention to this particular application problem and for providing access to the Chagas data base used. Work partially funded by grant MTM2013-43917-P from the Spanish Ministry of Science and Education, grant PAPIIT IN114814 of the Dirección General de Asuntos del Personal Académico of the Universidad Nacional Autónoma de México and grant CONACYT number 241195.

## References

1. A. Adin, M. A. Martínez-Beneito, P. Botella-Rocamora, T. Goicoa, M. D. Ugarte. (2016). Smoothing and high risk areas detection in space-time disease mapping: a comparison of P-splines, autoregressive, and moving average models. *Stochastic Environmental Research and Risk Assessment*. DOI 10.1007/s00477-016-1269-8
2. Blangiardo M., Cameletti M., Baio G. and Rue H. (2013). Spatial and Spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, **4**: 33-49.
3. Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society B*, **36**, 192-236.
4. Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1-21.
5. Besag, J., Green, P.J., Higdon, D. M. and Mengersen, K. L., (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3-66.
6. Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. . *Journal of the Royal Statistical Society: Serie B*, **61** (4): 691-746.
7. Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley and Sons.
8. Canale, D. M., Cecere, M. C., Chuit, R., and Guoertler, R. E. (2000). Peridomestic distribution of *Triatoma garciabesi* and *Triatoma guasayana* in northwest Argentina. *Medical and veterinary entomology*, 14(4), 383-390.
9. Chia-Hsien Lin, Tzai-Hung Wen , Hwa-Jen Teng, Niann-Tai Chang. (2014). The spatio-temporal characteristics of potential dengue risk assessed by *Aedes aegypti* and *Aedes albopictus* in high-epidemic areas. *Stochastic Environmental Research and Risk Assessment*. DOI 10.1007/s00477-014-0940-1
10. Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671-681.
11. Coura, J.R., Junqueira, A. C., Fernandes, O., Valente, S. A., and Miles, M. A. (2002). Emerging chagas disease in Amazonian Brazil. *Trends in parasitology*, 18(4), 171-176.
12. Cressie, N.A.C. (1993) *Statistics for Spatial Data (Revised ed.)*. Wiley, New York.
13. Cruz-Reyes A and Pickering-López J.M. (2006). Chagas disease in Mexico: an analysis of geographical distribution during the past 76 years - A review. *Mem. Inst. Oswaldo Cruz* **101**(4), 345-354.
14. De la Vega G, Medone P, Ceccarelli S, Rabinovich J and Schilman PE. (2015). *Geographical distribution, climatic variability and thermo-tolerance of Chagas disease vectors*. *Ecography* **38**, 1-10
15. Díaz-Avalos, C., Peterson, D.L., Alvarado, E., Ferguson, S.A. and Besag, J.E. (2001) Space-time modelling of lightning-caused ignitions in the Blue Mountains, Oregon. *Canadian Journal of Forest Research*, **31** , 1579-1593
16. Díaz-Avalos C. (2007). Spatial modeling of habitat preferences of biological species using markov random fields. *Journal of Applied Statistics*, **34** (7), 807-821.
17. Galvao, C., and Justi, S. A. (2015). An overview on the ecology of Triatominae (Hemiptera: Reduviidae). *Acta Tropica*, 1521: 116-125.
18. Gorla, D. E. (2002). Variables ambientales registradas por sensores remotos como indicadores de la distribución geográfica de *Triatoma infestans* (Heteroptera: Reduviidae). *Ecología austral*, 12(2), 117-127.
19. Gneiting, T., Kleiber, W. and Schlather, M. (2010). *Matérn Cross-Covariance functions for multivariate random fields*. *Journal of the American Statistical Association*. **105** (491), 1167-1177.
20. Gurgel-Goncalves R., Galvão C., Costa J., and Peterson A T. (2012). Geographic Distribution of Chagas Disease Vectors in Brazil Based on Ecological Niche Modeling. *Journal of Tropical Medicine*, 1-15.
21. Gurtler, R.E., and Yadon, Z. E. (2015). Eco-bio-social research on community-based approaches for Chagas disease vector control in Latin America. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 109(2), 91-98.
22. Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.
23. Harvill J. L. (2010). *Spatio-temporal processes*. *WIREs Computational Statistics*, **2**, 375-382.
24. Illian J. B. and Hendrichsen D. K.(2010). Gibbs point process models with mixed effects. *Environmetrics*. **21**: 341-353.
25. Illian J.B., Sorbye S.H. and Rue H.(2012). A toolbox for fitting complex spatial point processes models using integrated nested Laplace approximations (INLA). *The Annals of Applied Statistics*. **6** Issue 4, 1499-1530.
26. Illian, J.B., Martino, S., Sorbye, S.H., Gallego-Fernández, J.B., Zunzunegui, M., Esquivias, M.P. and Travis, J.M.J.(2013). Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution*, **4** (4): 305-315.
27. Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, **29**(4):597?614.
28. Knorr-Held L, Schrödle B, Rue H.(2009). Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA (2009), a chapter in statistical modelling and regression structures: *Festschrift in honour of Ludwig Fahrmeir*. In: Tutz G, Kneib T (eds) *Physica-Verlag, Heidelberg*. Available at <http://www.r-inla.org/papers>.



29. Lindgren F. and Rue H. (2013). Bayesian spatial and spatio-temporal modelling with R-INLA. *Journal of Statistical Software*.
30. Lindgren F, Rue H, and Lindstrom J (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach (with discussion). *Journal of the Royal Statistical Society, Series B*, **73** (4): 423-498.
31. Medone, P., Ceccarelli, S., Parham, P. E., Figuera, A., and Rabinovich, J. E. (2015). The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **370**: 20130560 .
32. Nouvellet, P., Cucunube, Z. M., and Gourbitre, S. (2015). Chapter Four-Ecology, Evolution and Control of Chagas Disease: A Century of Neglected Modelling and a Promising Future. *Advances in parasitology*, **87**, 135-191.
33. Peterson, A. T., Sánchez-Cordero, V., Beard, C. B., and Ramsey, J. M. (2002). Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerging infectious diseases*, **8**, 662-667.
34. Plummer M. (2008). Penalized loss functions for Bayesian model Comparison. *Biostatistics*. **9**(3), 523-539.
35. Qi, Y and Wu, J. (1996). Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices. *Landscape Ecology*, **11**(1), 39-49.
36. R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
37. R-INLA project, accessed on August 13, 2012.
38. Rue H. and Martino S. (2006). Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference*. **137**: 3177-3192.
39. Rue H, Martino S. and Chopin N. (2007). *Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations*. Statistics Report No. 1, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
40. Rue H, Martino S, Chopin N (2009). Approximate Bayesian Inference for Latent Gaussian Models using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society B*. **71**, 319-392.
41. Schofield, C. J., Jannin, J., and Salvatella, R. (2006). The future of Chagas disease control. *Trends in parasitology*, **22**(12), 583-588.
42. Simpson D, Illian J, Lindgren F, Sorbye S.H. and Rue H. (2011). *Going off grid: computationally efficient inference for log-Gaussian Cox processes*. NTNU Technical Report 10, Trondheim University.
43. Schmunis G.A. and Yadon Z.E. (2010). Chagas disease: A Latin American health problem becoming a world health problema. *Acta Tropica*, **115**: 14-21.
44. Taylor, B.M. and Diggle, P. (2014) INLA or MCMC?. A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*. **84**(10), 2266-2284.
45. Tierney L. and Kadane J.B. (1986). Accurate Approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*. **81**: 82-86.
46. Ugarte, M. D., Ibáñez, B. and Militino, A. F. (2005). Detection of spatial variation in risk when using CAR models for smoothing relative risks. *Stochastic Environmental Research and Risk Assessment*, **19** (1):33-40.
47. World Health Organization (2002). Control of Chagas disease. Second report of the WHO Expert Committee. W.H.O. Tech. Rep. Ser. 905, Geneva, 1-109.
48. Zeledon, R., and Rabinovich, J. E. (1981). Chagas disease: an ecological appraisal with special emphasis on its insect vectors. *Annual review of entomology*, **26**(1): 101-133.

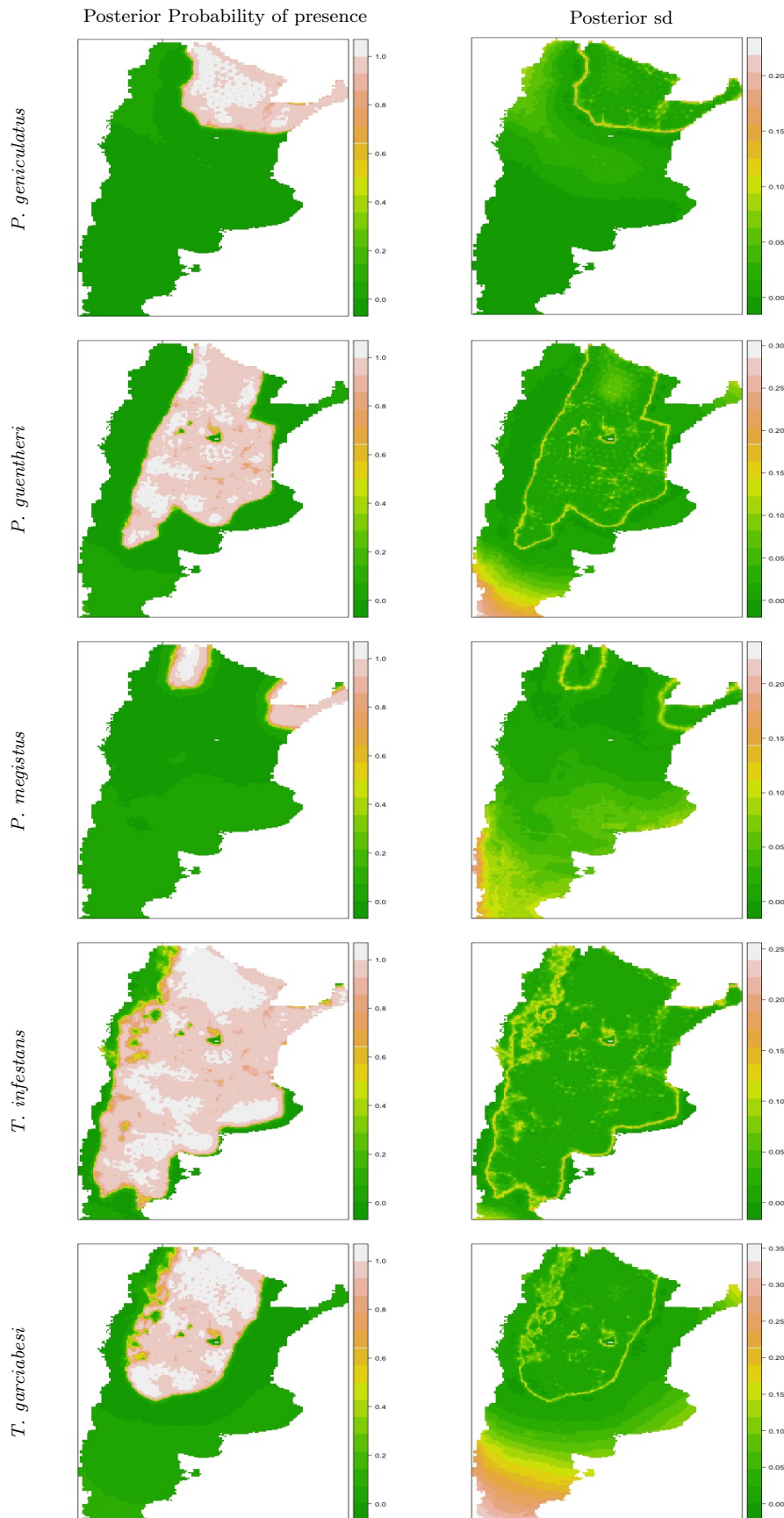
**Table 2.** Summary statistics for autologistic models fitted using MCMC: posterior median, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
<i>P. geniculatus</i>					
without covariates					
Intercept	-1773.65	380.5328	-2646.615	-1773.65	-1623.118
with covariates					
Intercept	-2.533	1.881	-4.901	-2.533	2.508
Altitude	-6.108	0.870	-6.587	-6.108	-3.806
Rain	1.326	0.715	0.656	1.326	3.008
Sunshine	-29.486	8.847	-30.513	-29.486	-2.716
AET	21.622	8.851	1.989	21.622	26.049
LST	13.912	6.046	1.308	13.912	17.084
Humidity	-32.028	11.908	-35.222	-32.028	-1.007
<i>P. guentheri</i>					
without covariates					
Intercept	-25.047	15.775	-54.280	-25.047	-0.254
with covariates					
Intercept	16.518	5.821	-0.592	16.51	17.092
Altitude	-2.237	1.044	-4.200	-2.237	-1.284
NDVI	10.628	3.733	0.388	10.628	15.692
AET	58.121	21.657	2.020	58.121	64.587
LST	26.807	7.692	3.568	26.807	27.595

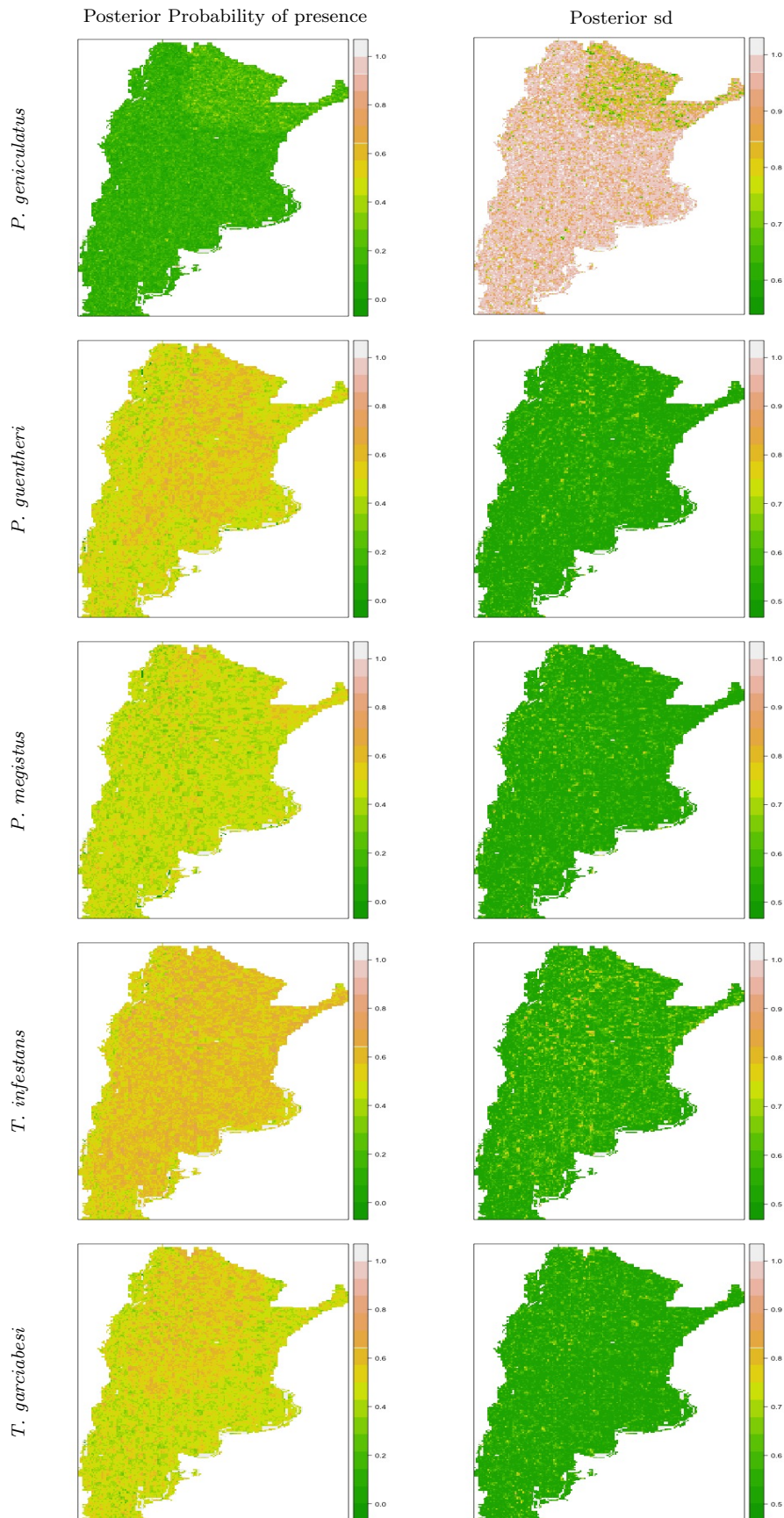
**Table 3.** Summary statistics for autologistic models fitted using MCMC: posterior median, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
<i>P. megistus</i>					
without covariates					
Intercept	-125.275	69.058	-232.485	-125.275	-5.141
with covariates					
Intercept	3.525	3.255	-3.719	3.525	7.549
Altitude	0.180	0.519	-1.384	0.180	0.579
NDVI	7.743	2.942	0.9744	7.743	9.984
Sunshine	-39.663	14.572	-53.692	-39.663	-1.838
LST	-18.661	4.918	-21.544	-18.661	-1.0167
Humidity	-50.601	19.671	-68.749	-50.601	-3.126
<i>T. Infestans</i>					
without covariates					
Intercept	82.879	43.191	7.000	82.879	146.979
with covariates					
Intercept	1.601	1.313	-0.939	1.601	4.021
Altitude	-0.738	0.171	-1.027	-0.738	-0.385
Sunshine	16.150	3.848	3.668	16.150	19.130
AET	12.844	4.519	2.582	12.844	19.311
LST	7.903	3.792	0.586	7.903	13.148
Humidity	7.217	1.949	3.195	7.217	10.993

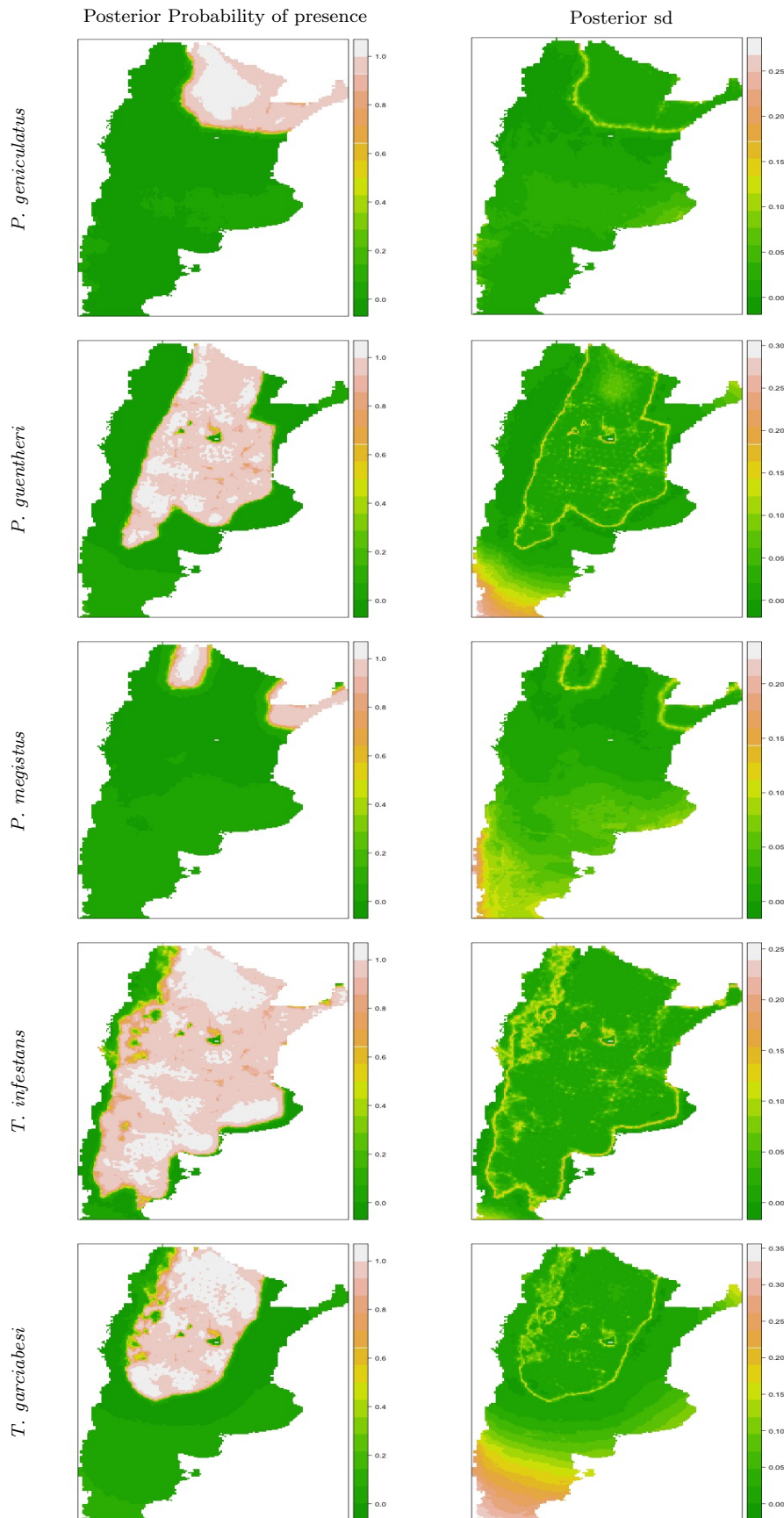
**6 APPENDIX**



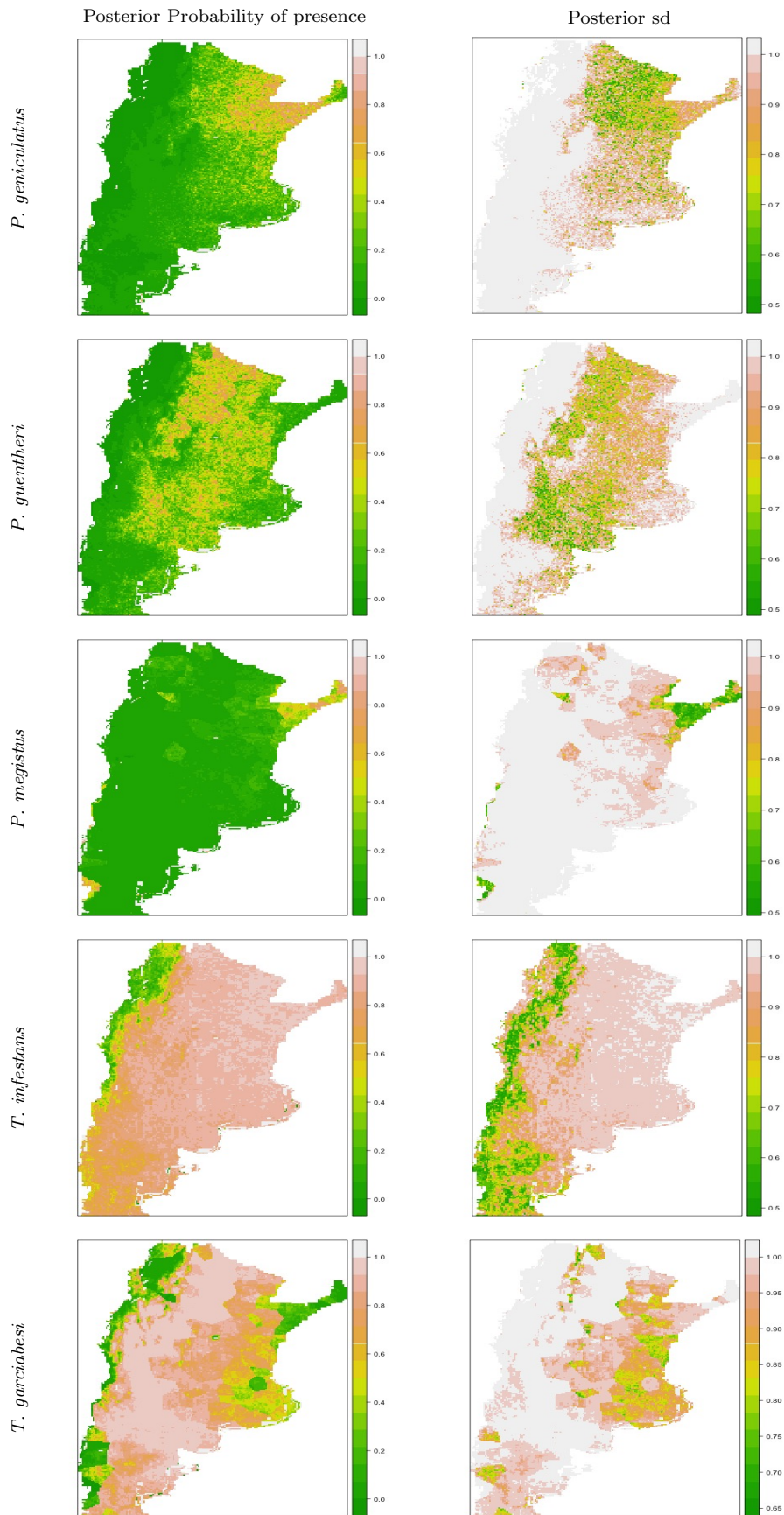
**Fig. 5.** Posterior probability of presence and standard deviation maps for autologistic models fitted without covariates using INLA-SPDE for the insect vector species considered in this study.



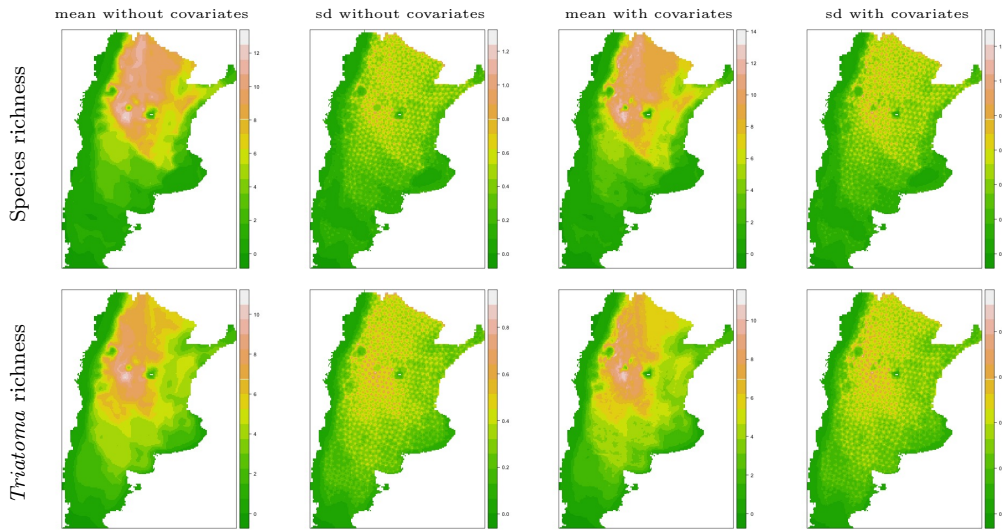
**Fig. 6.** Posterior probability of presence and standard deviation maps for autologistic models fitted without covariates using MCMC for the insect vector species considered in this study.



**Fig. 7.** Posterior probability of presence and standard deviation maps for autologistic models fitted with covariates using INLA-SPDE for the insect vector species considered in this study.



**Fig. 8.** Posterior probability of presence and standard deviation maps for autologistic models fitted with covariates using MCMC for the insect vector species considered in this study.



**Fig. 9.** Mean and sd maps for species richness models without covariates (first two columns from left to right) and with covariates (third and fourth columns) for the insect vector species considered in this study.

**Table 4.** Summary statistics for autologistic models fitted using MCMC: posterior mean, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
<i>T. garciabesi</i>					
without covariates					
Intercept	-65.231	35.049	-118.120	-65.231	-2.277
with covariates					
Intercept	-49.928	54.484	-134.281	-49.928	48.147
LST	124.380	61.103	10.218	124.38	206.03
Humidity	124.380	61.1036	10.2187	124.380	206.030

**Table 5.** Summary statistics for autologistic models fitted with INLA-SPDE: posterior mean, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
<i>P. geniculatus</i>					
without covariates					
Intercept	-25.357	3.717	-33.010	-25.235	-18.384
with covariates					
Intercept	-53.499	8.571	-70.778	-53.367	-36.933
NDVI	15.554	3.840	8.082	15.531	23.152
AET	0.2610	0.0817	0.1032	0.2599	0.4248
LST	0.5199	0.1239	0.2777	0.5195	0.7639
<i>P. guentheri</i>					
without covariates					
Intercept	-10.0769	30.3547	-67.7663	-10.7804	51.344
with covariates					
Intercept	-7.1593	30.5665	-67.1694	-7.1609	52.8052
Altitude	0.0020	0.0007	0.0007	0.0020	0.0034
NDVI	18.993	2.382	14.443	18.948	23.803
Rain	-0.012	0.004	-0.022	-0.012	-0.002
AET	0.427	0.113	0.208	0.426	0.652
LST	0.540	0.094	0.356	0.540	0.728
Aridity	0.322	0.122	0.083	0.322	0.564
Humidity	-0.067	0.027	-0.120	-0.067	-0.013

**Table 6.** Summary statistics for autologistic models fitted with INLA-SPDE: posterior mean, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
<i>P. megistus</i>					
without covariates					
Intercept	-37.911	4.070	-46.516	-37.694	-30.542
with covariates					
Intercept	-15.775	7.619	-30.703	-15.800	-0.730
NDVI	6.371	2.687	1.147	6.352	11.695
Sunshine	-0.031	0.013	-0.057	-0.031	-0.005
AET	0.179	0.061	0.060	0.179	0.299
Humidity	-0.062	0.024	-0.110	-0.062	-0.014
<i>T. infestans</i>					
without covariates					
Intercept	-10.112	15.702	-41.748	-10.417	24.453
with covariates					
Intercept	-17.123	18.234	-50.096	-18.739	26.355
NDVI	16.057	1.141	13.847	16.046	18.328
Humidity	0.021	0.008	0.006	0.021	0.037



**Table 7.** Summary statistics for autologistic models fitted with INLA-SPDE: posterior mean, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
<i>T. garciabesi</i>					
without covariates					
Intercept	-18.869	24.881	-67.716	-18.871	29.94
with covariates					
Intercept	-11.890	29.246	-68.712	-12.108	46.173
Altitude	0.0006	0.000	0.000	0.000	0.001
NDVI	18.7073	1.661	15.509	18.685	22.031

**Table 8.** Summary statistics for auto Poisson models fitted with INLA-SPDE: posterior mean, posterior standard deviation (SD) and posterior 95% credible interval for fixed effects

	Mean	SD	2.5%	50%	97.5%
Species richness					
without covariates					
Intercept	-3.188	12.492	-31.59	-3.466	27.517
with covariates					
Intercept	-4.473	12.594	-32.399	-4.926	27.093
NDVI	1.541	0.107	1.330	1.541	1.753
LST	0.023	0.003	0.017	0.023	0.030
<i>Triatoma</i> richness					
without covariates					
Intercept	-1.979	20.1408	-43.971	-2.445	42.169
with covariates					
Intercept	-2.788	20.346	-44.634	-3.462	42.162
Altitude	0.000	0.000	0.000	0.000	0.000
NDVI	1.601	0.105	1.395	1.601	1.808
LST	0.023	0.003	0.016	0.023	0.030