

# A multivariate geostatistical framework for combining multiple indices of abundance for disease vectors and reservoirs: A case study of *rattiness* in a low-income urban Brazilian community

Max T. Eyre<sup>1,2</sup>, Ticiana S. A. Carvalho-Pereira<sup>3</sup>, Fábio N. Souza<sup>3</sup>, Hussein Khalil<sup>3,4</sup>, Kathryn P. Hacker<sup>5</sup>, Soledad Serrano<sup>6</sup>, Joshua P. Taylor<sup>6</sup>, Mitermayer G. Reis<sup>3,7</sup>, Albert I. Ko<sup>7,8</sup>, Mike Begon<sup>9</sup>, Peter J. Diggle<sup>1</sup>, Federico Costa<sup>3,7,8,\*</sup>, and Emanuele Giorgi<sup>1,\*</sup>

<sup>1</sup>Centre for Health Informatics, Computing, and Statistics, Lancaster University Medical School, UK

<sup>2</sup>Liverpool School of Tropical Medicine, Liverpool, UK

<sup>3</sup>Institute of Collective Health, Federal University of Bahia, Salvador, Bahia, Brazil

<sup>4</sup>Swedish University of Agricultural Sciences, Umeå, Sweden

<sup>5</sup>University of Pennsylvania, Philadelphia, Pennsylvania, USA

<sup>6</sup>Instituto de Investigaciones Forestales y Agropecuarias Bariloche (IFAB), San Carlos de Bariloche, Río Negro, Argentina

<sup>7</sup>Oswaldo Cruz Foundation, Brazilian Ministry of Health, Salvador, Bahia, Brazil

<sup>8</sup>Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, USA

<sup>9</sup>Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, UK

\*These authors contributed equally to this work

July 31, 2020

## Abstract

A key requirement in studies of endemic vector-borne or zoonotic disease is an estimate of the spatial variation in vector or reservoir host abundance. For many vector species, multiple indices of abundance are available, but current approaches to choosing between or combining these indices do not fully exploit the potential inferential benefits that might accrue from modelling their joint spatial distribution. Here, we develop a class of multivariate generalized linear geostatistical models for multiple indices of abundance. We illustrate this novel methodology with a case study on Norway rats in a low-income urban Brazilian community, where rat abundance is a likely risk-factor for human leptospirosis. We combine three indices of rat abundance to draw predictive inferences on a spatially continuous latent process, *rattiness*, that acts as a proxy for abundance. We show how to explore the association between *rattiness* and spatially varying environmental factors, evaluate the relative importance of each of the three contributing indices, assess the presence of residual, unexplained spatial variation, and identify *rattiness* hotspots. The proposed methodology is applicable more generally as a tool for understanding the role of vector or reservoir host abundance in predicting spatial variation in the risk of human disease.

## 1 Introduction

In studies of endemic vector-borne and zoonotic diseases, estimates of vector and reservoir host abundance, including spatial variation in abundance, are often needed to inform predictive models of disease risk and to guide the decision-making process for the implementation, monitoring and evaluation of control programmes [1]. Detecting all members of a target population at a sampled location is impossible for most disease vector or reservoir species. Consequently, indirect methods of determination are often used in ecological studies to

obtain indices that quantify relative abundance [2–4]. Here, since our focus is on the effect of vector and reservoir host populations on human health, we use the term “abundance” loosely to denote all ecological processes that are associated with animal abundance, for example animal presence and activity, and that can be used to quantify exposure, including spatial variation in exposure, to a disease of interest.

In the absence of a gold-standard index of animal abundance, many different indices are commonly used for a single species, sometimes within the same study. For example, in the case of rodents, indices derived from traps, camera traps, counts (of animals, tracks, burrows and faeces), track plates and gnawing pegs have all been used to estimate rat abundance [2, 5, 6]. Similarly, for insects, a wide range of entomological indices are used. For example, occurrence, density (per unit time, surface area, person or other sampling unit), human-biting rates and the human blood index are used to estimate adult mosquito abundance [7–10]. When data for multiple imperfect indices of abundance has been collected within a study area, methods which can jointly model these quantities may improve prediction and inference. Such methods are also useful when different traps (or protocols) with different detection probabilities and biases are used to collect data for the same index [11, 12]. However, current approaches to combining multiple indices of abundance do not exploit the inferential benefits that might accrue from their joint modelling.

Many recent studies have attempted to model the spatial distribution of disease vectors, for example in the context of malaria [12], dengue [13], Chagas disease [14], Human African trypanosomiasis [15], schistosomiasis [16], leishmaniasis [17], West Nile virus [18] and rodent-borne zoonoses such as leptospirosis [19], plague [20], hantavirus [21] and *Bartonella* spp. [22]. In such cases, direct determination of vector or reservoir host abundance throughout the study area is often impractical, because of the extensive sampling effort required. A practical solution is to sample a finite set of locations and use statistical modelling to make predictions at unobserved locations by capturing spatial correlation and associations with environmental drivers of abundance. Here, we achieve this by the use of model-based geostatistics [23, 24], a branch of spatial statistics that provides a principled likelihood-based approach for mapping of geo-referenced outcomes. A geostatistical model is an extension of a generalised linear mixed model that accounts for covariate effects and otherwise unexplained spatial variation in the outcome of interest. Geostatistics has been used in a range of scientific disciplines, including ecology [25, 26] and epidemiology [24].

In this study, therefore, having described our motivating application in Section 1.1, in Section 2.1 we set out statistical criteria for combining multiple indices of vector and reservoir host abundance and review the literature for existing and relevant methodologies. Then, in Sections 2.2 to 2.4 we present a new class of multivariate generalized linear geostatistical models for combining multiple indices of abundance, which exploit the spatial correlation both within and across indices. In Section 3, we illustrate the development and application of the methodology in the context of a case study on the Norway rat, a reservoir for infectious diseases in low-income urban communities in Salvador, Brazil. Mapping Norway rat abundance is essential for investigating its role in disease transmission and developing more targeted rodent control strategies. In Section 4, we apply our novel methodology to the analysis of data collected for three indices of rat abundance, make inferences about the association of environmental variables with *rattiness*, our proxy for rat abundance, and map *rattiness* for the entire study area. We then assess the relative contribution of each index to the spatial predictions. Finally, in Section 5 we discuss the strengths, limitations and wider applicability of the developed methodology.

## 1.1 Motivating application: Mapping the abundance of Norway rats, a reservoir for *Leptospira* in urban Brazil

Leptospirosis is a widespread and neglected zoonotic disease caused by pathogenic bacteria from the *Leptospira* genus. It is among the leading zoonotic causes of morbidity and mortality globally, with more than one million human cases and 58,000 deaths reported each year [27, 28]. Humans are infected via direct contact with animal reservoirs or through contact with soil or water contaminated by bacteria shed in the urine of infected animals [29, 30]. In tropical low- and middle-income countries, including Brazil, low-income urban communities (often referred to as ‘informal settlements’ or ‘slum communities’ in the literature) are at increased risk for leptospirosis due to poor sanitary conditions, flooding, intense environmental contact and abundant local rat

reservoir populations [31, 32].

Globally, the Norway rat, *Rattus norvegicus*, is a major reservoir host for *Leptospira* spp. and several other pathogens, and thrives in low-income peri-urban and urban environments where food and harbourage are plentiful [30, 33–37]. Norway rat populations have been found to have high prevalences of leptospiral infection in Brazil, Argentina, Japan and Canada [19, 30, 36, 38], and high daily leptospire shedding rates have been recorded in Salvador, Brazil [30].

Associations between the risk of leptospiral infection in humans and peri-domiciliary rat infestation [31, 32, 34, 35, 39] and rodent sightings [31, 32, 35] have been reported in multiple studies in Brazil. However, the link between rat population abundance and risk of spill-over infection to humans is poorly understood [32], partly due to limited knowledge about the distribution and abundance of rats within urban environments [6]. As a result, several ongoing eco-epidemiological studies in Salvador, Brazil, aim to address this knowledge gap and generate evidence about the impact of rat control measures on disease transmission [40] through the collection and analysis of human seroprevalence and rat abundance data.

However, estimation of rat abundance in complex urban settings is hindered by a lack of reliable measurement tools [2, 6]. In studies of the Norway rat in Salvador, a combination of rat trapping, surveys for signs of rodent infestation and track plates are routinely used as indices of relative abundance [6, 41]. Track plates are plastic plates which are coated in ink and placed on the ground to detect rat paw and tail markings. A recent study has shown that track plate measurements are correlated with those of rodent infestation surveys and rat trapping [6]. The use of alternative tools, such as track plates, which are cheaper and can be deployed faster and more easily, allows for a more cost-effective design of studies, while reducing the impact of the loss of sampling days and equipment due to violence (associated with drug-trafficking groups operating within these communities) and theft. However, different indices may have distinct biases and collecting data for multiple indices within a study site can allow for a richer and more comprehensive measurement of relative abundance. In this challenging context, statistical modelling can be especially useful in order to use of all the information collected from multiple measurement tools and deliver optimized inferences about rat abundance.

Our research question is: how should we develop a joint geostatistical model for multiple indices to map rat abundance? Our ultimate goal is to develop a reliable modelling approach that can be used to identify rat abundance hotspots to guide future investigations of environmental contamination and rodent control.

## 2 Model and methods

### 2.1 Developing an approach for combining multiple indices of abundance

To develop an objective and statistically principled approach for combining multiple indices of abundance, we propose that a statistical model should meet the following six criteria.

- C1 It should account for the appropriate sampling distribution of each index through the use of a suitable conditional likelihood function.
- C2 It should not require all indices to be taken at a common set of locations.
- C3 It should account for spatial correlation both within and across indices.
- C4 It should allow for the prediction of abundance at all locations within the study area and quantify the uncertainty associated with those predictions.
- C5 It should allow for the quantification of the relative contribution of each index to the spatial predictions.
- C6 It should allow for the incorporation of spatially referenced covariates.

We now review existing methodologies in the literature and assess how well they meet these criteria.

One of the simplest and most commonly used approaches is to directly combine data for multiple indices into a single index by averaging their values and modelling this using a standard linear regression model [42]; this approach violates criteria C1 to C3. A second approach is to model each index separately and independently using linear regression models, and to combine the resulting predictions [42]. Although this approach respects C1, it does not take advantage of the inferential benefits that would accrue from C3.

In a third approach, a composite index is created using a weighted combination of multiple indices. The weighting is often based on a subjective theoretical framework derived from expert opinion [43]. Alternatively, summaries with specific weightings can be used, such as the General index [44] and the geoindex mean of multiple relative abundance indices (often used to quantify biodiversity [45]). A fourth approach is to obtain composite indices using Principal Component Analysis (PCA). This follows a more data-driven approach to combine multiple indices into a single real-valued score. These composite indices are commonly used for estimating general indicators of ecological systems [46], such as ecological integrity [47] and multispecies biodiversity indicators [48], rather than abundance. These methods do not respect any of the criteria C1 to C4, and C6 [43].

Geostatistical methods have been developed for modelling multiple indices of animal abundance for a single species. However, these methods were found to either use one index as a predictor for another index [49], thus violating C1 and C2, or to use multivariate kriging for all indices [50], violating C1 and C6.

There are several examples of geostatistical approaches which jointly model indices for multiple species in the field of ecological community modelling [42, 51, 52]. However, the structure of these models does not enable predictions to be made for the abundance of a single species measured by multiple indices, as is required for C4. They were also found to require all indices to be measured at a common set of locations, hence violating C2. While integrated species distribution models (ISDMs) offer a means to model multiple indices, they have been developed to combine multiple presence-only and presence-absence data sources [42, 53], rather than abundance indices. These models also provide no way to explore the relative contributions of each data source. Consequently, they do not meet C4 and C5.

## 2.2 Model formulation and inference

Let  $R(x)$  denote a spatially continuous stochastic process, representing *rattiness*, our proxy for rat abundance. The data consist of a set of outcomes  $Y_i = (Y_{i,j} : j = 1, \dots, J)$ , for  $i = 1, \dots, N$ , collected at a discrete set of locations  $X = \{x_i : i = 1, \dots, N\}$ . The outcome variables  $Y_j : j = 1, \dots, J$  are a set of indices that provide information about  $R(x)$ .

Let “[.]” be a shorthand notation for “the probability distribution of .” Define  $Y = (Y_1, \dots, Y_N)$  and  $R = (R(x_1), \dots, R(x_N))$ . We assume that the  $Y_{i,j} : j = 1, \dots, J$  are conditionally independent given  $R(x_i)$ , as formally expressed by

$$[Y|R] = \prod_{i=1}^N \prod_{j=1}^J [Y_{i,j}|R(x_i)]. \quad (1)$$

Let  $g_j(\cdot)$  and  $\eta_j(x_i)$  denote the link function and linear predictor for the outcome variables  $Y_{i,j} : i = 1, \dots, N$ . Hence,

$$\begin{aligned} g_j \{ \mu_j(x_i) \} &= \eta_j(x_i) = \alpha_j + \sigma_j R(x_i) \\ R(x_i) &= d^T(x_i) \beta + \sqrt{\psi} S(x_i) + \sqrt{1 - \psi} U_i \end{aligned} \quad (2)$$

where:  $d(x_i)$  is a vector of explanatory variables with associated regression coefficients  $\beta$ ;  $U_i$  is a set of independently and identically distributed zero-mean Gaussian variables with unit variance;  $S(x_i)$  is a stationary and isotropic spatial Gaussian process;  $\sigma_j > 0 : j = 1, \dots, J$  are scale parameters that account for the different scales of variation of the linear predictors of each outcome  $Y_{i,j}$ ;  $\psi \in (0, 1)$  regulates the relative contributions of spatially structured variation,  $S(x_i)$ , and unstructured random variation,  $U_i$ , to  $R(x_i)$ .

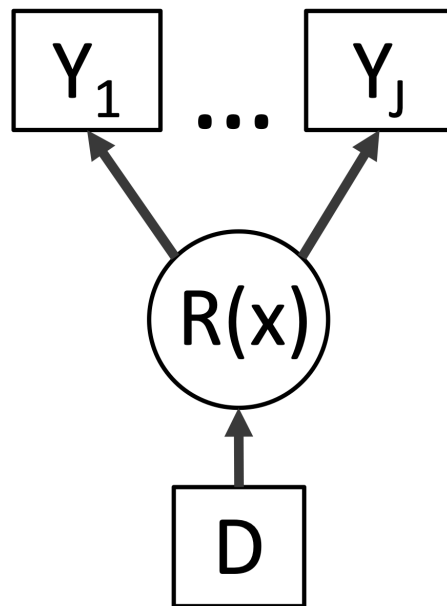


Figure 1: Directed acyclic graph (DAG) of the geostatistical model of Section 2.2.  $R(x)$  is the value of a spatially continuous stochastic process at location  $x$ . The outcome variables  $Y_j : j = 1, \dots, J$  are a set of indices that provide information about  $R(x)$ . The term  $D$  represents a set of explanatory variables which contribute to the spatial variation in  $R(x)$ . Square objects correspond to observable variables, and circles to latent random variables.

For the analysis of Section 3, we specify an exponential spatial correlation function

$$\text{Corr}\{S(x), S(x')\} = e^{-u/\phi}$$

where  $u = \|x - x'\|$  is the Euclidean distance between  $x$  and  $x'$ , and  $\phi$  regulates how fast the spatial correlation decays to zero with increasing distance  $u$ .

To fit the model in (2), we use the Monte Carlo maximum likelihood (MCML) method [54] and proceed as follows. Let  $\theta = (\alpha_1, \dots, \alpha_J, \sigma_1, \dots, \sigma_J)$  and  $\omega = (\beta, \phi, \psi)$  be the vector of unknown parameters associated with  $[R]$  and  $[Y|R]$ . The likelihood function is then given by

$$L(\theta, \omega) = [Y; \theta, \omega] = \int_{\mathbb{R}^N} [R; \omega][Y|R; \theta] dR \quad (3)$$

Since the integral in (3) cannot be solved analytically, we approximate it using Monte Carlo methods. Specifically, let  $\theta_0$  and  $\omega_0$  be our initial best guesses for  $\theta$  and  $\omega$ , respectively. Since  $[R; \omega][Y|R; \theta] \propto [R|Y; \omega]$  we re-write the integral in (3) using an importance sampling distribution  $[R; \omega_0][Y|R; \theta_0]$  to give

$$\begin{aligned} L(\theta, \omega) &\propto \int_{\mathbb{R}^N} \frac{[R; \omega][Y|R; \theta]}{[R; \omega_0][Y|R; \theta_0]} [R|Y; \theta_0, \omega_0] dR \\ &= E \left[ \frac{[R; \omega][Y|R; \theta]}{[R; \omega_0][Y|R; \theta_0]} \right], \end{aligned} \quad (4)$$

where the expectation is taken with respect to the distribution of  $[R|Y; \omega_0]$ .

Based on (4), we then approximate (3) with

$$L(\omega, \theta) \approx \frac{1}{B} \sum_{b=1}^B \frac{[r^{(b)}; \omega][Y|r^{(b)}; \theta]}{[r^{(b)}; \omega_0][Y|r^{(b)}; \theta_0]} \quad (5)$$

where  $r_{(b)}$  is the  $b$ -th sample from  $[R|Y; \omega_0, \theta_0]$ . To obtain the maximum likelihood estimate for  $\theta$  and  $\omega$ , we maximize (5) using numerical optimization. To simulate from  $[R|Y; \theta_0, \omega_0]$ , we use the Laplace sampling algorithm described in detail in [54] and [55].

To improve the approximation of the likelihood function, we also update our guesses  $\omega_0$  and  $\theta_0$  by plugging in the maximum likelihood estimate and re-iterate the maximization of (5) until convergence.

## 2.3 Exploratory analysis

In this section we outline several key steps in an exploratory analysis of the data to guide the model-building process.

The exploratory analysis serves three purposes: i) to explore the relationship between the latent *rattiness* process,  $R(x)$ , and the covariates  $d(x)$ ; ii) to test for the presence of residual spatial variation in  $R(x)$  unexplained by the covariates  $d(x)$ ; and iii) to assess if the data support the assumed stochastic dependence structure as represented by the causal arrows of Figure 1.

To pursue i), we first analyse the data using a simplified version of the model in (2) that does not assume spatial correlation and does not make use of any of the available covariates by setting  $\psi = 0$  and  $\beta = 0$ . Rattiness is consequently modelled purely as unstructured random variation, hence  $R(x_i) = U_i$ . Note that the likelihood associated with  $\theta = (\alpha_1, \dots, \alpha_J, \sigma_1, \dots, \sigma_J)$  and  $\omega = (\phi)$ , is now given by the product of  $N$  one-dimensional integrals,

$$L(\omega, \theta) = [Y; \omega, \theta] = \prod_{i=1}^N \int_{\mathbb{R}} [U_i; \theta] [Y_i | U_i; \theta] dU_i \quad (6)$$

To maximize the likelihood, each of the factors in the above product can then be approximated using numerical quadrature; we use a quasi Monte Carlo method whereby the integrals in (2) are drawn deterministically based on the Halton sequence of support points [56].

After fitting the model in (6), we then estimate  $U_i$  using its predictive expectation,

$$\hat{U}_i = E[U_i | Y_i] = \frac{\int_{\mathbb{R}} U_i [U_i; \omega] [U_i | R; \theta] dU_i}{\int_{\mathbb{R}} [U_i; \omega] [U_i | R; \theta] dU_i}.$$

To compute the  $\hat{U}_i$  we plug in the maximum likelihood estimates for  $\theta$  and  $\omega$ . To explore the functional form of the relationship between  $R(x_i)$  and the explanatory variables  $d(x_i)$ , we plot the  $\hat{U}_i$  against the values of  $d(x_i)$ . Based on the empirical relationship observed in the scatter-plots, we then introduce  $d(x_i)$  into the linear predictor for  $R(x_i)$ , leading to a first extension of the simplified model to

$$R(x_i) = \beta^\top d(x_i) + U_i. \quad (7)$$

We can now use the model in (7) to pursue the second objective, i.e. testing for residual spatial correlation. We then re-fit the likelihood (6), without setting  $\beta$  to zero, and re-compute  $\hat{U}_i$ , which now represent our provisional estimate for the residual variation in  $R(x_i)$  that is unexplained by  $d(x_i)$ . To check if the  $\hat{U}_i$  show evidence of spatial correlation, we then randomly permute locations  $x_i$  in the data while holding the  $U_i$  fixed and repeat this 10,000 times. For each of the permuted data-sets, we compute the empirical variogram based on the  $\hat{U}_i$  and use the resulting 10,000 variograms to compute 95% confidence intervals under the assumption that the  $\hat{U}_i$  are not spatially correlated. If the variogram computed from the original  $\hat{U}_i$  falls fully within the 95% band, we conclude that the  $\hat{U}_i$  do not show evidence of residual spatial correlation. If the variogram partly falls outside the 95% band, we conclude that there is evidence of spatial correlation and fit the geostatistical model defined in Section 2.2.

Finally, for the third objective, we test the null hypotheses  $H_0 : \sigma_j = 0$ , for  $j = 1, \dots, J$  using likelihood ratio tests. Note that, in this case, the conditions that give an asymptotic distribution of the likelihood ratio

test based on a  $\chi^2$  distribution are not met because the value 0 is on the boundary of the parametric space of  $\sigma_j$ ; following [57], we correct the nominal resulting p-values by multiplying them by 1/2.

## 2.4 Spatial prediction and assessment of the contribution of each index of abundance

Our predictive target is  $T(x) = d(x)^\top \beta + \sqrt{\psi}S(x)$  at a set of prediction locations,  $X^* = \{x_1^*, \dots, x_H^*\}$ . To predict  $T^* = (T(x_1^*), \dots, T(x_H^*))$ , we sample from its predictive distribution  $[T^*|Y]$  as follows. We first simulate from  $[R|Y; \theta, \omega]$  using the same sampling algorithm as for maximizing the likelihood in Section 2.2, with the parameters  $\theta$  and  $\omega$  fixed at their maximum likelihood estimates. After obtaining samples  $r_{(b)}$ ,  $b = 1, \dots, B$ , we then simulate from  $[T^*|r_{(b)}]$ , which follow a multivariate Gaussian distribution with mean and covariance matrix easily obtained from their joint Gaussian distribution  $[R, T^*]$ . From this, we obtain  $t_{(b)}(x_h^*)$ , for  $h = 1, \dots, H$  and  $b = 1, \dots, B$ , which are now samples drawn from  $[T^*|Y]$ . These can be used to compute any desired summary of the predictive distribution  $[T^*|Y]$ , such that the expectation

$$\hat{t}(x_h^*) = E[T(x_h^*)|Y] \approx \frac{1}{B} \sum_{b=1}^B t_{(b)}(x_h^*), h = 1, \dots, H$$

and the standard deviation

$$s(x_h^*) = \text{sd}[T(x_h^*)|Y] \approx \sqrt{\frac{1}{B} \sum_{b=1}^B (t_{(b)}(x_h^*) - \hat{t}(x_h^*))^2}, h = 1, \dots, H.$$

To assess the contribution of each index to the prediction target  $T^*$ , we then compare  $\hat{t}(x_h^*)$  and  $s(x_h^*)$  from the geostatistical model fitted using all indices with those obtained from  $J$  models, each of which excludes a single index.

Let  $\hat{t}_{-j}(x_h^*)$  and  $s_{-j}(x_h^*)$  denote the predictive mean and standard deviation obtained from excluding data for the  $j$ -th index; to summarize the discrepancy between this and the full model, we average the squared differences across all locations  $X^*$ , i.e.

$$SQ_M = \frac{1}{H} \sum_{h=1}^H (\hat{t}_{-j}(x_h^*) - \hat{t}(x_h^*))^2$$

and

$$SQ_{SD} = \frac{1}{H} \sum_{h=1}^H (s_{-j}(x_h^*) - s(x_h^*))^2.$$

## 3 Case study: Norway rats in Pau da Lima community

### 3.1 Study site and data collection

Pau da Lima (13°32'53.47" S; 38°43'51.10" W) is a low-income urban community with a high annual leptospiral infection rate of 35.4 (95% CI, 30.7–40.6) infection events per 1,000 annual follow-up events in the period 2003 to 2007 [32, 58]. It is located on the periphery of the city of Salvador in Northeast Brazil and has been a focus for leptospirosis research for over fifteen years [30, 32, 41]. The study site at Pau da Lima is characterised by three valleys (see Figure 2) with large elevation gradients, high population density, poverty, low levels of education and poor provision of sanitation and refuse collection services.

To describe the spatial variation in rat abundance within Pau da Lima, a cross-sectional study was conducted from October to December 2014. Rat trapping was carried out at 159 locations across the study area

with two traps deployed for 4 consecutive 24-hour trapping periods at each point (see Panti-May *et al.* [41]). After each 24-hour period, trapping success and closure of the trap without a rat caught inside were recorded. Track plates were used for two consecutive 24-hour periods at 415 locations [59] using the standardised protocol for placement and survey developed and validated previously [6] with five plates placed at each location in the shape of a ‘five’ on a die with 1 metre spacing between each plate. After each 24-hour period, plates were repainted and lost plates were recorded and replaced. A map of the study area and sampling locations for rat trapping and track plates is shown in Figure 2. A survey for signs of rat infestation (presence of trails, faecal droppings and active burrows), adapted from the Centers for Disease Control and Prevention [60] and validated in the study area [34], was also conducted at all locations at which traps or track plates were deployed. In our analysis, we consider the following environmental variables: elevation, distance to public refuse dumps and the proportion of land cover classified as vegetation within a 30 metre radius. Land cover data was created by classification of Digital Globe’s WorldView-2 satellite imagery (8 bands with resolution 0.5m by 0.5m resolution taken on February 17, 2013) [61] using a maximum likelihood supervised algorithm. This was validated with ground truthing data collected from 20 randomly selected sites of size 5m by 5m.

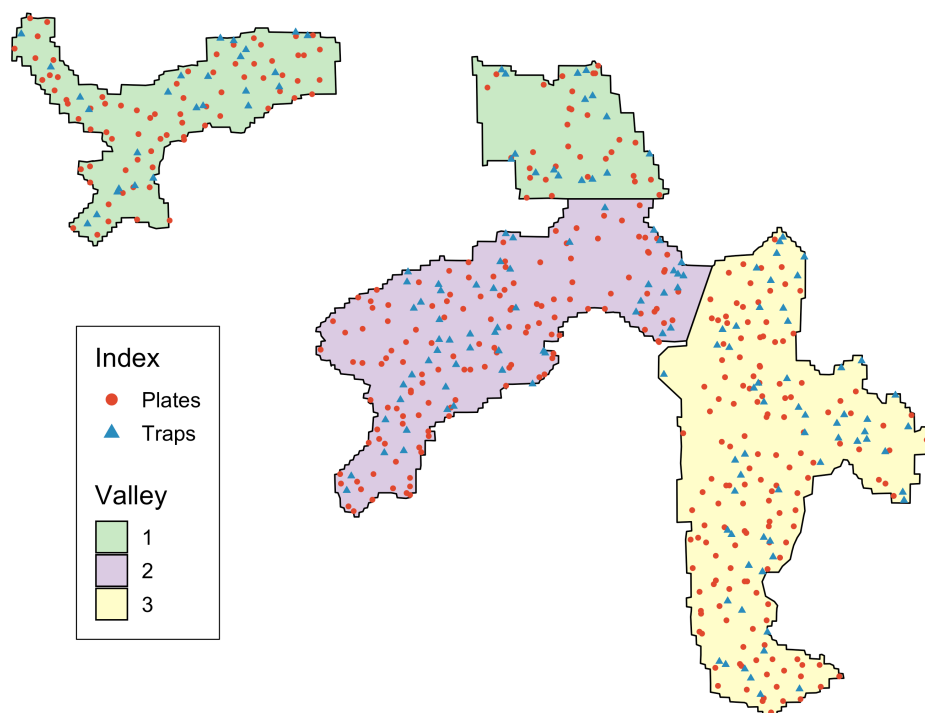


Figure 2: Map of the three valleys within the study site, Pau da Lima, with sampled locations for track plates and traps shown. Surveys for signs of rat infestation were conducted at all locations.

### 3.2 Applying the model to Pau da Lima data

In this analysis, we shall also refer to  $R(x)$  as the *rattiness* process. To make inference on  $R(x)$ , we shall use data collected on three indices: rat signs ( $j = 1$ ), live traps ( $j = 2$ ) and track plates ( $j = 3$ ).

Due to theft, violence and heavy rainfall, 993 out of 4150 (23.9%) plates were lost, with all five plates lost at 126 out of 830 (15.2%) track plate days. For the traps, 85 out of 1272 (6.7%) trapping-days were lost, of which 458 (38.6%) were found closed and empty after a 24-hour period. This is a common issue and is due to traps malfunctioning or being tampered with by animals or people. Of the remaining track plate-days, 263



(37.4%) had at least one plate with rat markings. Similarly, 200 (34.8%) of surveys found at least one sign of rat infestation. Of the trapping-days which were not lost, the trapping success rate was low, with only 112 (9.4%) trapping-days found to have caught a rat.

Let the variable  $Y_{i,1}$  be a binary indicator taking value 1, if at least one sign of rat infestation was found at location  $x_i$  and 0 otherwise. We model the probability of finding a sign of rat infestation,  $\mu_1(x_i)$ , using a logit-linear regression  $\log\{\mu_1(x_i)/(1 - \mu_1(x_i))\} = \alpha_1 + \sigma_1 R(x_i)$

The variable  $Y_{i,2}$ , conditionally on  $R(x_i)$ , is a Binomial variable representing the number of traps, out of  $n_{i,1}$ , in which rats were captured. We assume that the times of rat captures from a trap follow a time-varying inhomogeneous Poisson process with intensity  $t_i \mu_2(x_i)$ , where  $t_i$  is the time (in days) for which a trap is operative and  $\log\{\mu_2(x_i)\} = \alpha_2 + \sigma_2 R(x_i)$ . It follows that the probability of capturing a rat is

$$1 - \exp\{-t_i \mu_2(x_i)\}.$$

If a trap is found closed without a rat, we assume that the trap was disturbed and set  $t = 0.5$ . In all other cases, we set  $t = 1$ .

Finally,  $Y_{i,3}$ , is the number of track-plates, out of  $n_{i,3}$ , that show presence of rats. We model this as a Binomial variable with  $n_{i,3}$  trials and probability  $\mu_3(x_i)$  where  $\log\{\mu_3(x_i)/(1 - \mu_3(x_i))\} = \alpha_3 + \sigma_3 R(x_i)$ .

## 4 Results

### 4.1 Exploratory analysis

Following the steps outlined in Section 2.3, we first fit the simplified non-spatial model ( $R(x_i) = U_i$ ) to explore the relationship between *rattiness* and each of the environmental variables considered. The results are reported in the scatter-plots of Figure 3. In Figure 3a, we observe a negative linear relationship between elevation and *rattiness*. In Figure 3b, we notice that, on average, higher values of *rattiness* are observed for distances from dumps less than 90 meters. Finally, Figure 3c, shows that the mean proportion of land cover classified as vegetation (within a 30 meters radius of a given sampling point), is approximately linearly and positively associated with *rattiness*.

Based on the results of this exploratory analysis, we then extend the non-spatial model to include the covariate effects on *rattiness*. Hence, our model for  $R(x_i)$  becomes

$$R(x_i) = \sum_{l=1}^6 \beta_l d_l(x_i) + U_i \quad (8)$$

where each of the terms  $\beta_l$  corresponds to covariate effects as defined in Table 1,  $d_l$  are the explanatory variables and  $U_i$  are the unstructured random effects.

Table 1: Regression coefficients of the environmental covariates used to model *rattiness*.

Term	Variable
$\beta_1$	Elevation (m)
$\beta_2$	Refuse distance (0-90m)
$\beta_3$	Refuse distance (>90m)
$\beta_4$	Proportion of land cover vegetation (%)
$\beta_5$	Valley 2 (relative to Valley 1)
$\beta_6$	Valley 3 (relative to Valley 1)

We then carry out the likelihood ratio tests for testing the three hypothesis  $H_0 : \sigma_j = 0$  for  $j = 1, 2, 3$ . All three yield p-values less than 0.0001, supporting the use of a joint model for all three indices.

Figure 3d shows the variogram for the  $\hat{U}_i$  obtained from the spatially uncorrelated model (8). Most of the points of the empirical variogram lie inside the 95% tolerance band but the variogram point at around 10 meters is a highly unlikely value under the assumption of spatial independence. Since the variogram diagnostic does not provide an unequivocal answer to the question of whether a spatially correlated term is needed, we fit a geostatistical model to assess this.

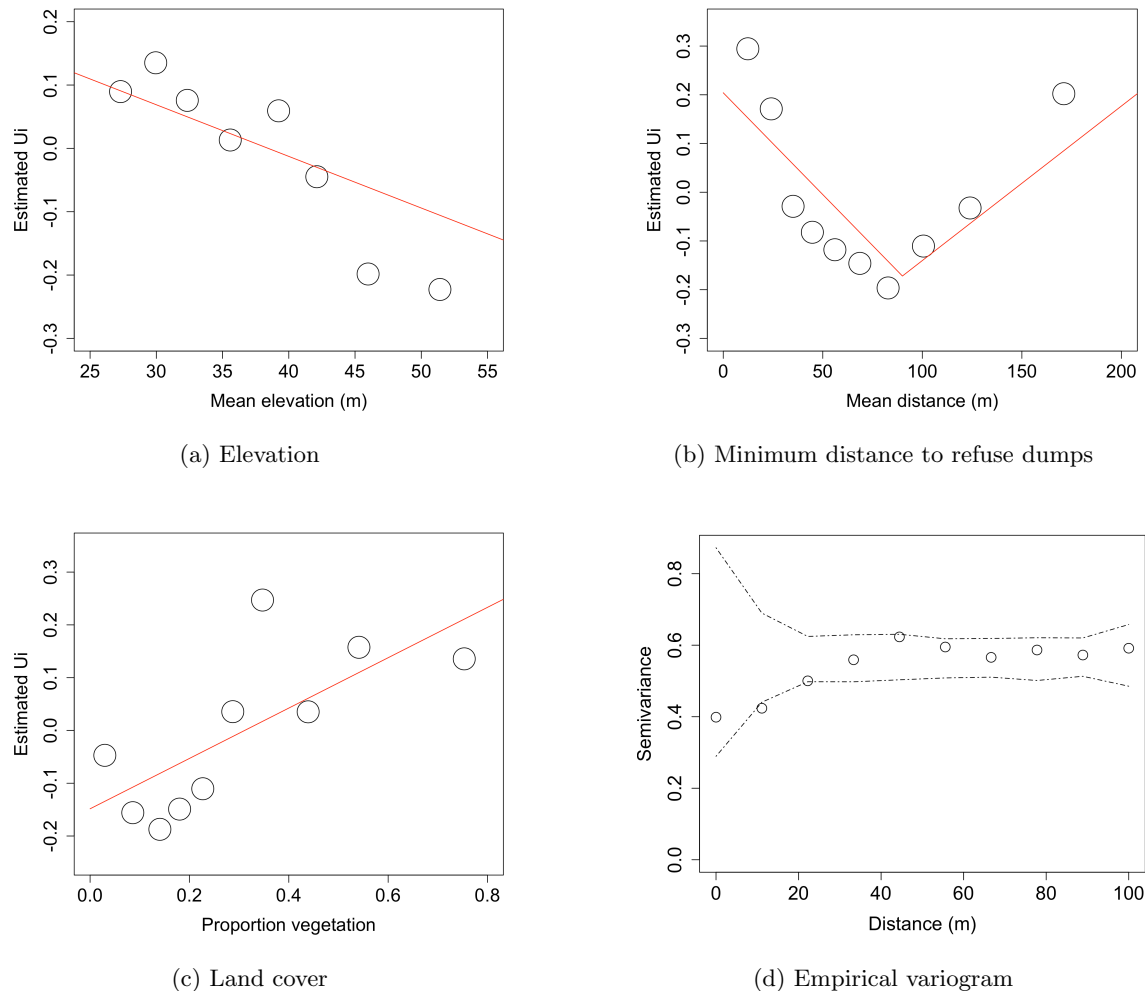


Figure 3: Panels (a) to (c) are scatter-plots of the unstructured random variation,  $\hat{U}_i$ , against the covariates considered in the analysis.  $\hat{U}_i$  are estimated using a non-spatial model which excludes all covariates. The relationship between each covariate and  $\hat{U}_i$  is shown with a red line and was estimated using univariable linear regression (with a linear piecewise spline knot included at 90 metres for the distance to refuse covariate). Panel (d) is a variogram computed from  $\hat{U}_i$  using a non-spatial model that includes all of the covariates; the dashed lines correspond to 95% confidence intervals under the assumption of spatial independence.

## 4.2 Geostatistical model

The parameter estimates based on the Monte Carlo maximum likelihood method are reported in Table 2 under the ‘full model’ column. The estimate for the scale of spatial correlation,  $\phi$ , of about 13 meters indicates that the data exhibit spatial correlation after controlling for the explanatory variables. The estimate for  $\psi$  of about 0.9 implies a greater contribution to *rattiness* from the Gaussian process,  $S(x_i)$ , than from the unstructured random variation,  $U_i$ , suggesting that most of the unexplained variation in *rattiness* is spatially structured. All

the point estimates of the regression coefficients  $\beta_l$  are consistent with the scatter-plots of Figure 3. Valleys 2 and 3 both had lower mean levels of *rattiness* relative to Valley 1, controlling for all other covariates. While the association with valley was significant at  $p < 0.05$ , elevation, distance to refuse dumps or land cover covariates were not significant.

Table 2: Parameter estimates for the full model and the three two-indices models where  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  (and  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$ ) denote the coefficients for Signs, Traps and Plates, respectively

Parameter	Estimate (95%CI)			
	Full model	Signs & Traps	Traps & Plates	Signs & Plates
$\alpha_1$	-0.642 (-0.891, -0.425)	-0.508 (-0.997, -0.115)	-	-0.630 (-0.853, -0.419)
$\alpha_2$	-2.684 (-3.078, -2.361)	-2.607 (-3.021, -2.099)	-2.456 (-2.849, -2.127)	-
$\alpha_3$	-2.503 (-2.925, -2.144)	-	-2.567 (-3.066, -2.119)	-2.672 (-3.090, -2.305)
$\sigma_1$	0.747 (0.455, 1.045)	1.040 (0.615, 1.356)	-	0.641 (0.372, 0.906)
$\sigma_2$	0.920 (0.613, 1.183)	1.068 (0.546, 1.182)	0.863 (0.436, 1.133)	-
$\sigma_3$	1.896 (1.557, 2.179)	-	1.795 (1.348, 2.174)	1.877 (1.542, 2.163)
$\phi$	13.432 (6.833, 21.172)	46.413 (7.692, 162.455)	11.306 (4.978, 17.325)	12.270 (6.920, 16.099)
$\psi$	0.878 (0.529, 1.000)	0.492 (0.161, 0.878)	-	-
$\beta_1$	-0.131 (-0.333, 0.058)	-0.337 (-0.795, -0.050)	-0.103 (-0.396, 0.158)	-0.164 (-0.370, 0.024)
$\beta_2$	-0.234 (-0.582, 0.101)	-0.173 (-0.819, 0.407)	-0.299 (-0.795, 0.175)	-0.055 (-0.401, 0.295)
$\beta_3$	-0.074 (-0.399, 0.260)	0.016 (-0.343, 0.354)	-0.129 (-0.493, 0.077)	-0.011 (-0.125, 0.248)
$\beta_4$	0.114 (-0.075, 0.310)	0.158 (-0.170, 0.527)	0.086 (-0.194, 0.376)	0.115 (-0.070, 0.308)
$\beta_5$	-0.229 (-0.358, -0.108)	0.116 (-0.170, 0.354)	-0.269 (-0.485, -0.085)	-0.361 (-0.503, -0.241)
$\beta_6$	-0.159 (-0.289, -0.034)	0.195 (-0.064, 0.459)	-0.220 (-0.401, -0.040)	-0.182 (-0.307, -0.059)

### 4.3 Spatial prediction

Our predictive target for *rattiness* is

$$T(x) = \sum_{l=1}^6 \beta_l d_l(x) + \sqrt{\psi} S(x)$$

for prediction locations  $x$  forming a 5 m by 5 m regular grid covering the whole of the study area.

Maps for the mean and standard deviation of the predictive distribution of  $T(x)$  are shown in Figure 4. These show a highly heterogeneous spatial pattern with localized hotspots in Valley 1 (Figure 4a) and in the low elevation central regions of the two other valleys. Areas with low values of predicted *rattiness* in Valleys 2 and 3 are characterized by a high proportion of soil land cover and higher elevations. The Gaussian process,  $S(x)$ , contributed more than the covariates to the prediction of the *rattiness* surface shown in Figure 4b. This is evidenced by the clear geographical overlap of hotspots in both  $S(x)$  and mean predicted *rattiness*. As expected, in areas with fewer or no observations, standard errors are larger than elsewhere.

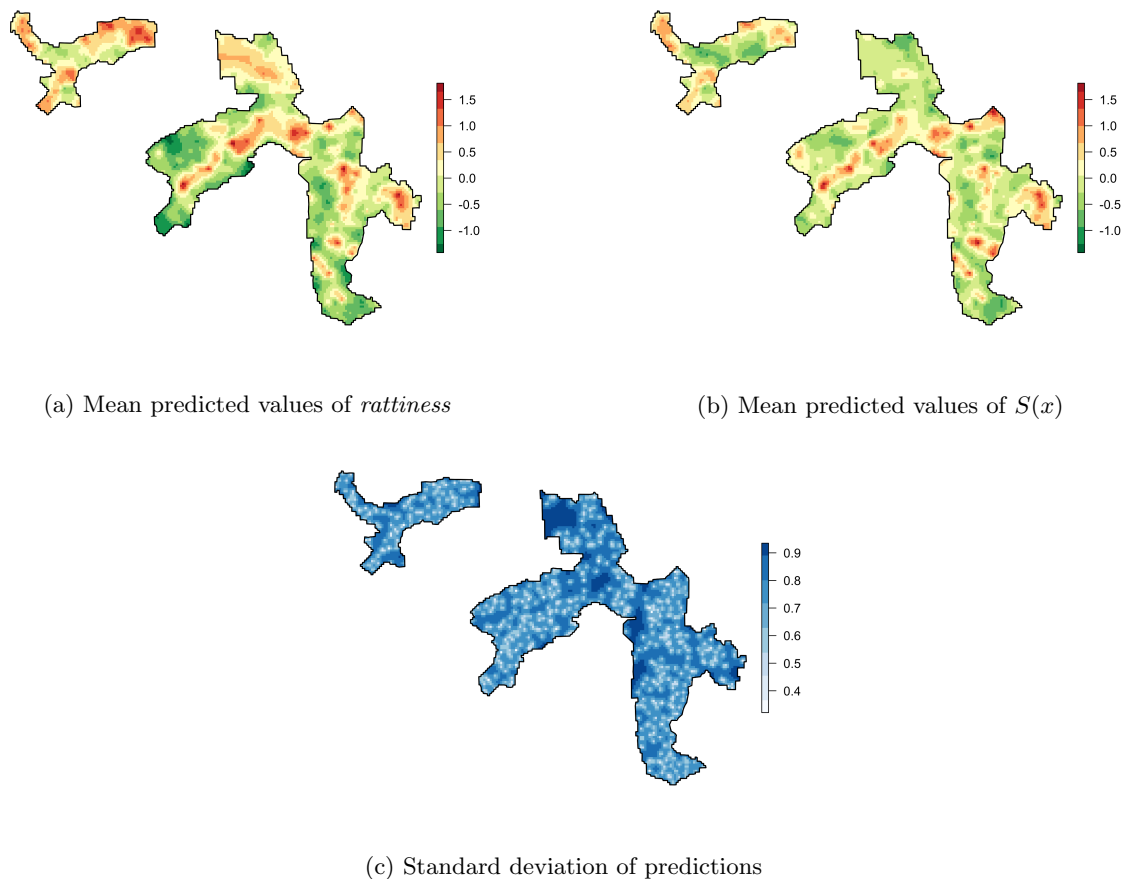


Figure 4: *Rattiness* model predictions (positive *rattiness* values indicate areas of high relative abundance)

#### 4.4 Relative contributions of indices

To assess the impact of each index on parameter estimation, we fitted three models, each discarding one of the three indices; the parameter estimates are reported in Table 2, and are similar to those estimated for the full model. Across all four models, the signs of covariate estimates were consistent except for the valley indicator variables, which vary both in sign and magnitude. The estimates for the scale of the spatial correlation,  $\phi$ , from the “Traps & Plates” and “Signs & Plates” models are close to that of the full model, but the “Signs & Traps” model had a substantially larger estimate of about 46 metres with a much wider confidence interval. The increased uncertainty in the estimation of the spatial correlation after excluding plates suggests that this index may be one of the main factors driving our predictions for *rattiness*. Furthermore, when the plates are included in the model, the spatial variation entirely dominates the *rattiness* process with estimates for  $\psi$  very close to 1. We therefore fixed  $\psi = 1$  for the “Track & Plates” and “Sign & Plates” models.

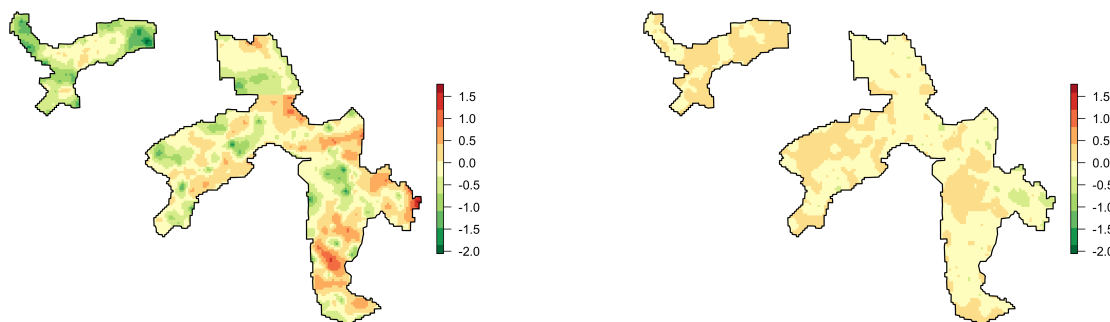
To visualize the differences in the spatial predictions for *rattiness* and  $S(x)$ , we compute the relative difference between the predictions obtained from each of the models excluding one of the three indices and the full model. Figure 5 reports the maps of the relative differences for *rattiness*, and Figure 6 for  $S(x)$ . The spatial predictions from the “Traps & Plates” and “Signs & Plates” models were more similar to those obtained from the full model as indicated by relative differences close to zero throughout the study area in both figures. In contrast, the predictions for the “Signs & Traps” model were different to those made by the three other models in most parts of the study area, with relative differences ranging from about  $-2$  to  $+1$ .

Table 3 reports the  $SQ_M$  and  $SQ_{SD}$  summaries used to quantify the changes in point predictions and

standard deviations for *rattiness*; see Section 2.4 for a formal definition of these two summaries. The results clearly highlight the model excluding plates as yielding substantially different predictions for *rattiness*, as well as larger standard errors. These results are consistent with our findings from Table 2, providing further evidence on the importance of plates to our predictive inferences on *rattiness*.

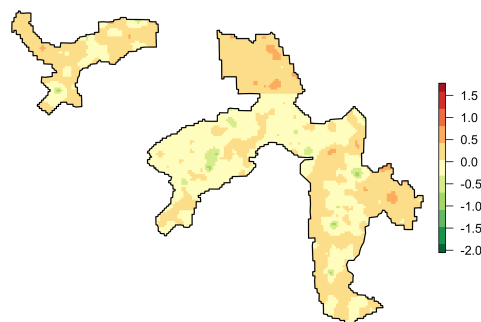
Table 3: The squared differences of the point predictions ( $SQ_M$ ) and standard deviations ( $SQ_{SD}$ ) between the full model and each of the three two-indices models, averaged over all the prediction locations. For a formal definition of  $SQ_M$  and  $SQ_{SD}$ , see Section 2.4.

Model	$SQ_M$	$SQ_{SD}$
Traps & Plates	0.018	0.007
Signs & Plates	0.032	0.003
Signs & Traps	0.249	0.096



(a) Signs and traps model

(b) Traps and plates model



(c) Signs and plates model

Figure 5: Model predictions for *rattiness* for each two-indices model relative to the full model (a value of zero indicates that there is no difference in the models' predicted values)

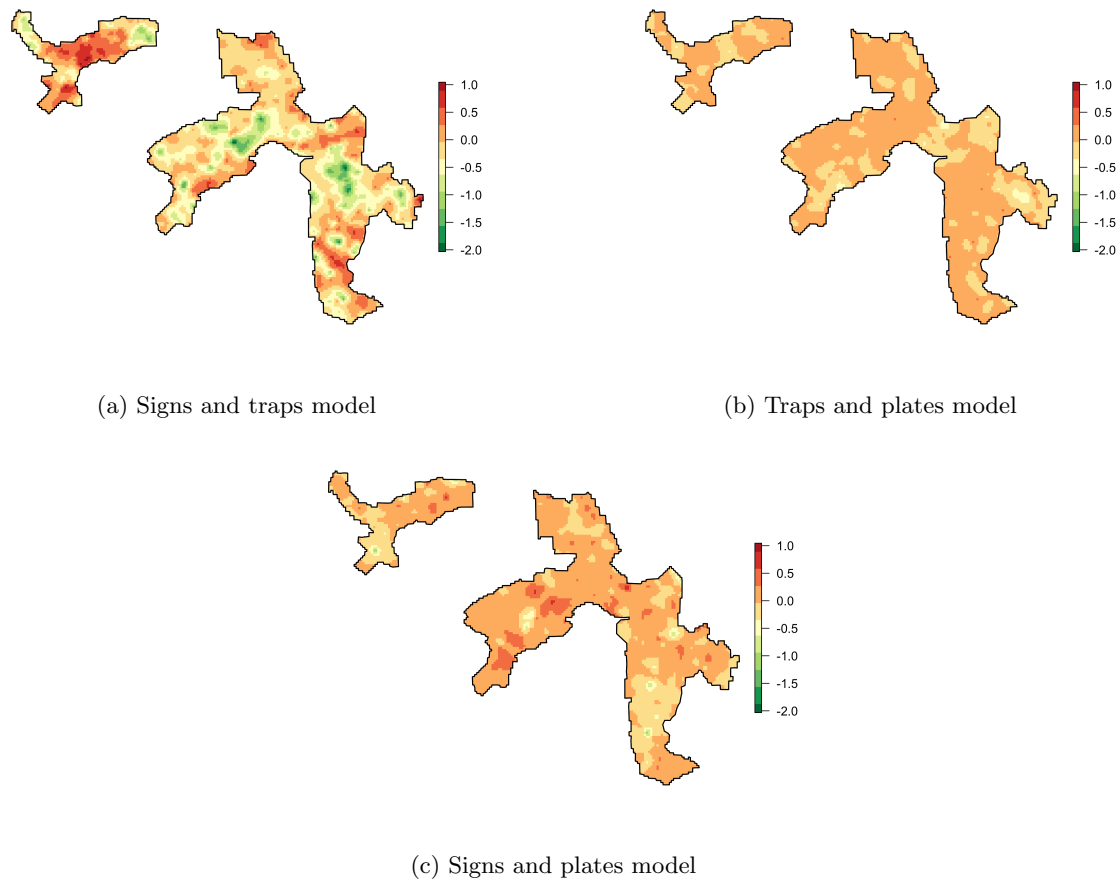


Figure 6: Model predictions for the spatial Gaussian process,  $S(x)$ , for each two-indices model relative to the full model (a value of zero indicates that there is no difference in the models' predicted  $S(x)$  values)

## 5 Discussion

In this paper, we have developed a flexible geostatistical framework that borrows information across multiple indices of vector and reservoir host abundance to carry out spatial prediction for a shared latent variable that acts as a proxy for animal abundance. To our knowledge this is the first study that proposes a multivariate geostatistical framework to jointly model multiple indices of abundance for a single species using a statistically principled likelihood-based approach.

We have applied the method to mapping *rattiness*, a proxy for Norway rat abundance, in a low-income urban community in Salvador, Brazil. We found that *rattiness* is lower at higher elevations and longer distances from large refuse piles, and higher in more densely vegetated areas. In our study site, elevation was used as a proxy for socioeconomic status within the community, with improved housing quality, sanitary conditions and road surfacing found at higher elevations. The point estimates for the regression coefficients associated with each of these variables are consistent with previous studies [37, 62], and can be explained by Norway rats' preference for habitats with greater access to food and harbourage. However, the inherently high sampling variation in the data recorded by each of the three indices results in wide confidence intervals for these estimates. A separate analysis that included elevation as the only covariate for *rattiness* (Supplementary Material 1) supports the conclusion that this was a key source of uncertainty; the effect of elevation was statistically significant at the conventional 5% level. Measurement error in covariates is also a likely contributor to these wide confidence intervals. This is a common issue when prediction is the priority, as was the case in our application, and

covariates are constrained to those for which values are available at all prediction locations.

A strength of our model is its ability to borrow information across space without requiring data from multiple indices to be co-located. This is especially useful when combining data for multiple indices from separate studies, or when there is a non-negligible loss or malfunctioning of measurement tools [63]. This arose in our analysis of the Norway rat due to track-plate loss and empty closing of traps, and is commonly encountered when measuring rodent abundance [2]. The framework also provides flexibility for modelling indices of abundance that are measured on a wide range of different scales with different sampling distributions. This was useful in our application as we were able to use a non-standard approach to more accurately model the trapping process and account for trap failure.

The estimate for the spatial correlation parameter,  $\hat{\phi} \approx 13$  metres corresponds to a spatial correlation range (the distance at which the correlation reduces to 0.05) of approximately 40 metres. This is consistent with studies investigating the size of the main activity area of Norway rats, which has been estimated to have a radius of 25–150m in urban areas [64, 65]. In environmentally heterogeneous and resource-rich areas, such as Pau da Lima, the size of a rat’s activity space has been found to be smaller, as shown by estimates of population density varying significantly within a city block [66] or along the length of an alley [67]. This is due to strong spatial heterogeneity in the presence of food and harbourage, availability of access routes and the presence of barriers to movement. All of these can result in high site fidelity (a measure of how concentrated an animal’s movements are around a specific site) [66] and significant variations in the abundance and activity of rats over small distances. The estimated value of  $\psi$  indicates that the spatially structured random effects are more important than the non-structured random effects in predicting *rattiness*. This follows from the fact that they account for unmeasured variables of habitat suitability, which can be expected to be spatially structured.

The finding that track plates were an important contributor for *rattiness* estimation indicates the greater information content provided by this tool relative to traps and signs. Nevertheless, the other two indices also contributed to *rattiness*, enabling more precise predictive inferences to be made than could be obtained using only the track-plates data. Due to the scarcity of resources available for monitoring programmes of vector and animal reservoir populations, efficient data collection is critical, making the choice of which indices, or how many should be used, an important consideration. A key strength of our methodology is that it provides the user with the tools to explore the contributions of each index to *rattiness* (or the spatial latent process for any other application). The likelihood-ratio tests described in Section 2.3 can be used to check which indices contribute to the model for prediction of *rattiness*. Any two such indices must both be associated with the latent *rattiness* variable,  $R(x)$ , in Figure 1, and therefore necessarily with each other. However, if two indices are near-collinear the likelihood-ratio test will indicate that one of the two is redundant. The  $SQ_M$  and  $SQ_{SD}$  summaries then enable the user to evaluate the extent to which each included index contributes to the predictions.

The interpretation of the *rattiness* process is an important, context-dependent issue. For example, if all the indices used in the analysis are reliable indicators of abundance, then *rattiness* has a clear interpretation as an overall measure of abundance. However, tools used to estimate abundance often provide measures that are a mixture of both animal abundance and behaviour [3]. In our application, the rat signs are an index of rat presence, while track plates and traps both measure abundance and activity, with track plates more strongly representative of activity. The resulting measure of *rattiness* obtained by combining the three indices therefore represents a data-determined synthesis of these three processes. In our motivating example concerning the role of rats in determining risk of human leptospirosis, the mechanism through which the vector confers disease risk is driven by both the size of the rat population and its behaviour, of which activity is one aspect [68, 69]. For this reason, we argue that joint modelling of multiple indices can be especially relevant to understanding geographical variation in disease risk.

The proposed modelling framework can also be applied to problems of environmental management and conservation, where indices of abundance are widely used to monitor wild animal populations and their impact on biodiversity, agriculture or another species’ ability to survive, and to guide management decisions [3, 63, 70]. Examples of recent studies that have used multiple indices of abundance include: invasive roof rat and deer mouse populations in orchards [71]; the threatened survival of native species due to invasive small mammals

on islands [72]; the impact of rats and possums in New Zealand [73]; and, the effect of elephants on woody vegetation in sub-Saharan Africa [74].

One limitation of our approach is that it does not account for detection bias. Methods that account for this bias require absolute abundance data, which is difficult to collect, or data collected using double-sampling techniques that require an often unattainable trapping success rate - for rodents, of at least 20% [2, 75]. For this reason, index data is still widely collected for monitoring purposes without the requirement for these additional data sources.

## References

- [1] Kitron U. Landscape Ecology and Epidemiology of Vector-Borne Diseases: Tools for Spatial Analysis. *Journal of Medical Entomology*. 1998;35(4):435–445. Available from: <https://doi.org/10.1093/jmedent/35.4.435>.
- [2] Cavia R, Rubn G, Virginia O. Techniques to Estimate Abundance and Monitoring Rodent Pests in Urban Environments. *Integrated Pest Management and Pest Control - Current and Future Tactics*. 2012;.
- [3] Stephens PA, Pettorelli N, Barlow J, Whittingham MJ, Cadotte MW. Management by proxy? The use of indices in applied ecology. *Journal of Applied Ecology*. 2015;52(1):1–6.
- [4] Falcy MR, McCormick JL, Miller SA. Proxies in practice: Calibration and validation of multiple indices of animal abundance. *Journal of Fish and Wildlife Management*. 2016;7(1):117–128.
- [5] O’Connell AF, Nichols JD, Karanth KU. *Camera traps in animal ecology: methods and analyses*. Springer Science & Business Media; 2010.
- [6] Hacker KP, Minter A, Begon M, Diggle PJ, Serrano S, Reis MG, et al. A comparative assessment of track plates to quantify fine scale variations in the relative abundance of Norway rats in urban slums. *Urban Ecosystems*. 2016;19(2):561–575.
- [7] Silver JB. *Mosquito ecology: field sampling methods*. springer science & business media; 2007.
- [8] WHO. *Malaria surveillance, monitoring & evaluation: a reference manual*; 2018. Available from: <http://www.who.int/malaria/publications/atoz/9789241565578/en/>.
- [9] Brown HE, Paladini M, Cook RA, Kline D, Barnard D, Fish D. Effectiveness of Mosquito Traps in Measuring Species Abundance and Composition. *Journal of Medical Entomology*. 2008;45(3):517–521.
- [10] Boyer S, Foray C, Dehecq JS. Spatial and temporal heterogeneities of *Aedes albopictus* density in La Reunion Island: Rise and weakness of entomological indices. *PLoS ONE*. 2014;9(3):1–12.
- [11] Tangena JAA, Thammavong P, Hiscox A, Lindsay SW, Brey PT. The human-baited double net trap: An alternative to human landing catches for collecting outdoor biting mosquitoes in Lao PDR. *PLoS ONE*. 2015;10(9):1–13.
- [12] Sinka ME, Golding N, Massey NC, Wiebe A, Huang Z, Hay SI, et al. Modelling the relative abundance of the primary African vectors of malaria before and after the implementation of indoor, insecticide-based vector control. *Malaria Journal*. 2016;15(1):1–10.
- [13] Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *elife*. 2015;4:e08347.
- [14] Dumonteil E, Gourbière S. Predicting *Triatoma dimidiata* abundance and infection rate: A risk map for natural transmission of Chagas disease in the Yucatán Peninsula of Mexico. *American Journal of Tropical Medicine and Hygiene*. 2004;70(5):514–519.



- [15] Stanton MC, Esterhuizen J, Tirados I, Betts H, Torr SJ. The development of high resolution maps of tsetse abundance to guide interventions against human African trypanosomiasis in northern Uganda. *Parasites and Vectors*. 2018;11(1):1–12.
- [16] Scholte RGC, Carvalho OS, Malone JB, Utzinger J, Vounatsou P. Spatial distribution of *biomphalaria* spp., the intermediate host snails of *schistosoma mansoni*, in Brazil. *Geospatial Health*. 2012;6(3 SUPPL.):95–101.
- [17] Ferro C, López M, Fuya P, Lugo L, Cordovez JM, González C. Spatial distribution of sand fly vectors and eco-epidemiology of cutaneous leishmaniasis transmission in Colombia. *PLoS ONE*. 2015;10(10):1–16.
- [18] Winters AM, Bolling BG, Beaty BJ, Blair CD, Eisen RJ, Meyer AM, et al. Combining mosquito vector and human disease data for improved assessment of spatial west Nile virus disease risk. *American Journal of Tropical Medicine and Hygiene*. 2008;78(4):654–665.
- [19] Himswoth CG, Bidulka J, Parsons KL, Feng AYT, Tang P, Jardine CM, et al. Ecology of *Leptospira interrogans* in Norway Rats (*Rattus norvegicus*) in an Inner-City Neighborhood of Vancouver, Canada. *PLoS Neglected Tropical Diseases*. 2013;7(6).
- [20] Wilschut LI, Addink EA, Heesterbeek JAP, Dubyanskiy VM, Davis SA, Laudoit A, et al. Mapping the distribution of the main host for plague in a complex landscape in Kazakhstan: an object-based approach using SPOT-5 XS, landsat 7 ETM+, SRTM and multiple random forests. *International Journal of Applied Earth Observation and Geoinformation*. 2013;23:81–94.
- [21] Carbajo AE, Pardiñas UFJ. Spatial Distribution Model of a Hantavirus Reservoir, the Long-tailed Ccolilargo (*Oligoryzomys longicaudatus*), in Argentina. *Journal of Mammalogy*. 2007;88(6):1555–1568.
- [22] Morand S, Bordes F, Blasdel K, Pilosof S, Cornu JF, Chaisiri K, et al. Assessing the distribution of disease-bearing rodents in human-modified tropical landscapes. *Journal of Applied Ecology*. 2015;52(3):784–794.
- [23] Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1998;47(3):299–350.
- [24] Diggle PJ, Giorgi E. *Model-based Geostatistics for Global Public Health: Methods and Applications*. Chapman and Hall/CRC; 2019.
- [25] Royle JA, Kéry M, Gautier R, Schmid H. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs*. 2007;77(3):465–481.
- [26] Rivoirard J, Simmonds J, Foote KG, Fernandes P, Bez N. *Geostatistics for estimating fish abundance*. John Wiley & Sons; 2008.
- [27] Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, et al. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Neglected Tropical Diseases*. 2015;9(9):0–1. Available from: <http://dx.doi.org/10.1371/journal.pntd.0003898>.
- [28] Torgerson PR, Hagan JE, Costa F, Calcagno J, Kane M, Martinez-Silveira MS, et al. Global Burden of Leptospirosis: Estimated in Terms of Disability Adjusted Life Years. *PLoS Neglected Tropical Diseases*. 2015;9(10):1–14.
- [29] Haake DA, Levett PN. Leptospirosis in humans. In: *Leptospira and leptospirosis*. Springer; 2015. p. 65–97.
- [30] Costa F, Wunder EA, de Oliveira D, Bisht V, Rodrigues G, Reis MG, et al. Patterns in *Leptospira* shedding in Norway rats (*Rattus norvegicus*) from Brazilian slum communities at high risk of disease transmission. *PLoS Neglected Tropical Diseases*. 2015;9(6):1–14. Available from: <http://dx.doi.org/10.1371/journal.pntd.0003819>.

- [31] Reis RB, Ribeiro GS, Felzemburgh RDM, Santana FS, Mohr S, Melendez AXTO, et al. Impact of environment and social gradient on *Leptospira* infection in urban slums. *PLoS Neglected Tropical Diseases*. 2008;2(4):11–18.
- [32] Hagan JE, Moraga P, Costa F, Capian N, Ribeiro GS, Wunder EA, et al. Spatiotemporal Determinants of Urban Leptospirosis Transmission: Four-Year Prospective Cohort Study of Slum Residents in Brazil. *PLoS Neglected Tropical Diseases*. 2016;10(1):1–17.
- [33] Mwachui MA, Crump L, Hartskeerl R, Zinsstag J, Hattendorf J. Environmental and Behavioural Determinants of Leptospirosis Transmission: A Systematic Review. *PLoS Neglected Tropical Diseases*. 2015;9(9):1–15.
- [34] Costa F, Ribeiro GS, Felzemburgh RDM, Santos N, Reis RB, Santos AC, et al. Influence of Household Rat Infestation on *Leptospira* Transmission in the Urban Slum Environment. *PLoS Neglected Tropical Diseases*. 2014;8(12).
- [35] Sarkar U, Nascimento S, Barbosa R, For MbCCIoRF, Leptospirosis During an Urban Epidemics R, Nuevo H, et al. Population-based case-control investigation of risk factors for leptospirosis during an urban epidemic. *American Journal Tropical Medicine and Hygiene*. 2002;66(5):605–10.
- [36] Vanasco NB, Sequeira MD, Sequeira G, Tarabla HD. Associations between leptospiral infection and seropositivity in rodents and environmental characteristics in Argentina. *Preventive Veterinary Medicine*. 2003;60(3):227–235.
- [37] Santos NdJ, Sousa E, Reis MG, Ko AI, Costa F. Rat infestation associated with environmental deficiencies in an urban slum community with high risk of leptospirosis transmission. *Cadernos de Saúde Pública*. 2017;33(2):1–13. Available from: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-311X2017000205003&lng=en&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2017000205003&lng=en&tlng=en).
- [38] Koizumi N, Muto M, Tanikawa T, Mizutani H, Sohmura Y, Hayashi E, et al. Human leptospirosis cases and the prevalence of rats harbouring *Leptospira interrogans* in urban areas of Tokyo, Japan. *Journal of Medical Microbiology*. 2009;58(9):1227–1230.
- [39] Maciel EAP, de Carvalho ALF, Nascimento SF, de Matos RB, Gouveia EL, Reis MG, et al. Household transmission of *Leptospira* infection in urban slum communities. *PLoS Neglected Tropical Diseases*. 2008;2(1).
- [40] Minter A, Diggle PJ, Costa F, Childs J, Ko AI, Begon M. A model for leptospire dynamics and control in the Norway rat (*Rattus norvegicus*) the reservoir host in urban slum environments. *Epidemics*. 2018;25(September 2017):26–34. Available from: <https://doi.org/10.1016/j.epidem.2018.05.002>.
- [41] Panti-May JA, Carvalho-Pereira TSA, Serrano S, Pedra GG, Taylor J, Pertile AC, et al. A Two-Year Ecological Study of Norway Rats (*Rattus norvegicus*) in a Brazilian Urban Slum. *Plos One*. 2016;11(3):e0152511. Available from: <http://dx.plos.org/10.1371/journal.pone.0152511>.
- [42] Fletcher RJ, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM. A practical guide for combining data to model species distributions. *Ecology*. 2019;100(6):1–15.
- [43] Dobbie MJ, Dail D. Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicators*. 2013;29:270–277. Available from: <http://dx.doi.org/10.1016/j.ecolind.2012.12.025>.
- [44] Engeman RM. Indexing principles and a widely applicable paradigm for indexing animal populations. *Wildlife Research*. 2005;32(3):203–210.
- [45] Buckland ST, Studeny AC, Magurran AE, Illian JB, Newson SE. The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere*. 2011;2(9):art100.
- [46] Heink U, Kowarik I. What are indicators? On the definition of indicators in ecology and environmental planning. *Ecological Indicators*. 2010;10(3):584–593.

- [47] Reza MIH. Importance and considerations for the development of a composite index of ecological integrity for ecological management. *International Journal of Ecology and Development*. 2014;28(2):32–48.
- [48] Vačkář D, Ten Brink B, Loh J, Baillie JEM, Reyers B. Review of multispecies indices for monitoring human impacts on biodiversity. *Ecological Indicators*. 2012;17:58–67.
- [49] Dey S, Delampady M, Parameshwaran R, Kumar NS, Srivathsa A, Karanth KU. Bayesian Methods for Estimating Animal Abundance at Large Spatial Scales Using Data from Multiple Sources. *Journal of Agricultural, Biological, and Environmental Statistics*. 2017;22(2):111–139.
- [50] Amiri K, Shabanipour N, Eagderi S. Using kriging and co-kriging to predict distributional areas of Kilka species (*Clupeonella* spp.) in the southern Caspian Sea. *International Journal of Aquatic Biology*. 2017;5(2):108–113.
- [51] Wang Y, Naumann U, Wright ST, Warton DI. Mvabund- an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution*. 2012;3(3):471–474.
- [52] Ovaskainen O, Soininen J. Making more out of sparse data: Hierarchical modeling of species communities. *Ecology*. 2011;92(2):289–295.
- [53] Pacifici K, Reich BJ, Miller DAW, Pease BS. Resolving misaligned spatial data with integrated species distribution models. *Ecology*. 2019;100(6):1–15.
- [54] Christensen OF. Monte Carlo maximum likelihood in model-based geostatistics. *Journal of computational and graphical statistics*. 2004;13(3):702–718.
- [55] Giorgi E, Diggle PJ. PrevMap: An R Package for Prevalence Mapping. *Journal of Statistical Software*. 2017;78(8):1–29.
- [56] Caffisch RE. Monte carlo and quasi-monte carlo methods. *Acta numerica*. 1998;7:1–49.
- [57] Self SG, Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*. 1987;82(398):605–610.
- [58] Hacker KP, Sacramento GA, Cruz JS, de Oliveira D, Nery N, Lindow JC, et al. Influence of Rainfall on *Leptospira* Infection and Disease in a Tropical Urban Setting, Brazil . *Emerging Infectious Diseases*. 2020;26(2):311–314.
- [59] Souza FN, Hacker KP, Eyre MT, Lobo C, Aguiar-Santos M, Carvalho-Pereira T, et al. Predicting the variation in the temporal and spatial presence of urban rats in tropical slums. (Unpublished);
- [60] for Disease Control C, Prevention, Others. Integrated pest management: conducting urban rodent surveys. Atlanta: US Department of Health and Human Services. 2006;.
- [61] Hacker KP, Minter A, Almeida C, Reinalde R, Begon M, Diggle PJ, et al. Use of tracking plates to identify hotspots of rat abundance in slum communities with high endemic transmission of leptospirosis. In: *Proceedings of the 65th Annual Meeting of the American Society of Tropical Medicine and Hygiene*. vol. 95. Atlanta, USA; 2016. p. 202.
- [62] De Masi E, Vilaça P, Razzolini MTP. Environmental conditions and rodent infestation in Campo Limpo district, São Paulo municipality, Brazil. *International Journal of Environmental Health Research*. 2009;19(1):1–16.
- [63] Engeman, Richard M, Witmer, Gary W. IPM strategies: Indexing difficult to monitor populations of pest species. *Proceedings of the Vertebrate Pest Conference*. 2000;19.
- [64] Traweger D, Travnitzky R, Moser C, Walzer C, Bernatzky G. Habitat preferences and distribution of the brown rat (*Rattus norvegicus* Berk.) in the city of Salzburg (Austria): implications for an urban rat management. *Journal of Pest Science*. 2006;79(3):113–125. Available from: <http://link.springer.com/10.1007/s10340-006-0123-z>.

- [65] Feng AYT, Himsworth CG. The secret life of the city rat: A review of the ecology of urban Norway and black rats (*Rattus norvegicus* and *Rattus rattus*). *Urban Ecosystems*. 2014;17(1):149–162.
- [66] Gardner-Santana LC, Norris DE, Fornadel CM, Hinson ER, Klein SL, Glass GE. Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (*Rattus norvegicus*). *Molecular ecology*. 2009;18(13):2766–2778.
- [67] Himsworth CG, Jardine CM, Parsons KL, Feng AYT, Patrick DM. The characteristics of wild rat (*Rattus* spp.) populations from an inner-city neighborhood with a focus on factors critical to the understanding of rat-associated zoonoses. *PLoS ONE*. 2014;9(3).
- [68] Hoverman JT, Searle CL. Behavioural influences on disease risk: implications for conservation and management. *Animal Behaviour*. 2016;120:263–271. Available from: <http://dx.doi.org/10.1016/j.anbehav.2016.05.013>.
- [69] Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to zoonotic spillover. *Nature Reviews Microbiology*. 2017;15(8):502–510. Available from: <http://dx.doi.org/10.1038/nrmicro.2017.45>.
- [70] Hopkins HL, Kennedy ML. An assessment of indices of relative and absolute abundance for monitoring populations of small mammals. *Wildlife Society Bulletin*. 2004;32(4):1289–1296.
- [71] Baldwin RA, Quinn N, Davis DH, Engeman RM. Effectiveness of rodenticides for managing invasive roof rats and native deer mice in orchards. *Environmental Science and Pollution Research*. 2014;21(9):5795–5802.
- [72] Towns DR, Atkinson IAE, Daugherty CH. Have the harmful effects of introduced rats on islands been exaggerated? *Biological Invasions*. 2006;8(4):863–891.
- [73] Ruffell J, Innes J, Didham RK. Efficacy of chew-track-card indices of rat and possum abundance across widely varying pest densities. *New Zealand Journal of Ecology*. 2015;39(1):87–92.
- [74] Nkosi SE, Adam E, Barrett AS, Brown LR. A synopsis of field and remote sensing based methods for studying African elephant (*Loxodonta Africana*) impact on woody vegetation in Africa. *Applied Ecology and Environmental Research*. 2019;17(2):4045–4066.
- [75] Sollmann R, Mohamed A, Samejima H, Wilting A. Risky business or simple solution - Relative abundance indices from camera-trapping. *Biological Conservation*. 2013;159:405–412. Available from: <http://dx.doi.org/10.1016/j.biocon.2012.12.025>.