INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Eyre MT *et al*. 2020 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. *J. R. Soc. Interface* **17**: 20200398. http://dx.doi.org/10.1098/rsif.2020.0398

Received: 27 May 2020
Accepted: 6 August 2020

Subject Category:
Life Sciences–Mathematics interface

Subject Areas:
environmental science

Keywords:
epidemiology, abundance indices, zoonotic and vector-borne diseases, multivariate model-based geostatistics, leptospirosis, Norway rat

Author for correspondence:
Max T. Eyre
e-mail: max.eyre@lstmed.ac.uk

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[1,2], Ticiana S. A. Carvalho-Pereira[3], Fábio N. Souza[3], Hussein Khalil[3,4], Kathryn P. Hacker[5], Soledad Serrano[6], Joshua P. Taylor[6], Mitermayer G. Reis[3,7], Albert I. Ko[7,8], Mike Begon[9], Peter J. Diggle[1], Federico Costa[3,7,8,†] and Emanuele Giorgi[1,†]

[1]Centre for Health Informatics, Computing, and Statistics, Lancaster University Medical School, Lancaster LA1 4YW, UK
[2]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK
[3]Institute of Collective Health, Federal University of Bahia, Salvador 40110-040, Bahia, Brazil
[4]Swedish University of Agricultural Sciences, Umeå 901 87, Sweden
[5]University of Pennsylvania, Philadelphia, PA 19104, USA
[6]Instituto de Investigaciones Forestales y Agropecuarias Bariloche (IFAB), Modesta Victoria 4450, 8400 San Carlos de Bariloche, Río Negro, Argentina
[7]Oswaldo Cruz Foundation, Brazilian Ministry of Health, Salvador 40296-710, Bahia, Brazil
[8]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT 06510, USA
[9]Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool L69 7ZB, UK

MTE, 0000-0001-9847-8632; KPH, 0000-0001-5267-1699; MB, 0000-0003-1715-5327; EG, 0000-0003-0640-181X

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 and 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 abundances, 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.

[†]These authors contributed equally to this work.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5096322.

THE ROYAL SOCIETY PUBLISHING

Consequently, indirect methods of determination are often used in ecological studies to obtain indices that quantify relative abundance [2–4]. Here, as 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 have been collected within a study area, methods that 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 generalized linearmixed 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 §1.1, in §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 §§2.2–2.4, 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 §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 §4, we apply our novel methodology to the analysis of data collected for three indices of rat abundance, which 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 §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 bacteria of the genus *Leptospira*. It is among the leading zoonotic causes of morbidity and mortality globally, with more than 1 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 an increased risk for leptospirosis owing 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 prevalence of leptospiral infection in Brazil, Argentina, Japan and Canada [19,30,36,38], and high daily *Leptospira* shedding rates have been recorded in Salvador, Brazil [30].

Association between the risk of leptospiral infection in humans and peri-domiciliary rat infestation [31,32,34,35,39] and rodent sightings [31,32,35] has 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 owing to the limited knowledge about the distribution and the 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 through the collection and analysis of human seroprevalence and rat abundance data [40].

However, estimation of rat abundance in complex urban settings is hindered by a lack of reliable measurement tools [2,6]. In studies on 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 that 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 owing 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 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–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. 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–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 that 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 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 $Y_i = (Y_{i,j} : j = 1, \ldots, J)$, for $i = 1, \ldots, N$, collected at a discrete set of locations $X = \{x_i : i = 1, \ldots, N\}$. The outcome variables $Y_j : j = 1, \ldots, J$ are a set of indices that provide information about $R(x)$.

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

$$[Y|R] = \prod_{i=1}^{N} \prod_{j=1}^{J} [Y_{i,j}|R(x_i)]. \tag{2.1}$$

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

$$\begin{aligned} g_j\{\mu_j(x_i)\} &= \eta_j(x_i) = \alpha_j + \sigma_j R(x_i) \\ R(x_i) &= d^{\mathsf{T}}(x_i)\beta + \sqrt{\psi}\, S(x_i) + \sqrt{1-\psi}\, U_i \end{aligned}, \tag{2.2}$$

where $d(x_i)$ is a vector of explanatory variables with associated regression coefficients β; 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; $\sigma_j > 0 : j = 1, \ldots, J$ 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 $R(x_i)$.

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

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

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

To fit the model in equation (2.2), we use the Monte Carlo maximum likelihood method [54] and proceed as follows. Let $\theta = (\alpha_1, \ldots, \alpha_J, \sigma_1, \ldots, \sigma_J)$ and $\omega = (\beta, \phi, \psi)$ be the vector of

unknown parameters associated with $[R]$ and $[Y \mid R]$. The likelihood function is then given by

$$L(\theta, \omega) = [Y; \theta, \omega] = \int_{\mathbb{R}^N} [R; \omega][Y|R; \theta]\, \mathrm{d}R. \tag{2.3}$$

Because the integral in equation (2.3) cannot be solved analytically, we approximate it using Monte Carlo methods. Specifically, let θ_0 and ω_0 be our initial best guesses for θ and ω, respectively. Because $[R; \omega][Y \mid R; \theta] \propto [R \mid Y; \omega]$, we re-write the integral in (2.3) using an importance sampling distribution $[R; \omega_0][Y \mid R; \theta_0]$ to give

$$\begin{aligned} L(\theta, \omega) &\propto \int_{\mathbb{R}^N} \frac{[R; \omega][Y|R; \theta]}{[R; \omega_0][Y|R; \theta_0]} [R|Y; \theta_0, \omega_0]\, \mathrm{d}R \\ &= E\left[\frac{[R; \omega][Y|R; \theta]}{[R; \omega_0][Y|R; \theta_0]}\right], \end{aligned} \tag{2.4}$$

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

On the basis of equation (2.4), we then approximate equation (2.3) with

$$L(\omega, \theta) \approx \frac{1}{B} \sum_{b=1}^{B} \frac{[r_{(b)}; \omega][Y|r_{(b)}; \theta]}{[r_{(b)}; \omega_0][Y|r_{(b)}; \theta_0]}, \tag{2.5}$$

where $r_{(b)}$ is the bth sample from $[R \mid Y; \omega_0, \theta_0]$. To obtain the maximum likelihood estimate for θ and ω, we maximize equation (2.5) using numerical optimization. To simulate from $[R \mid Y; \theta_0, \omega_0]$, we use the Laplace sampling algorithm described in detail in [54,55].

To improve the approximation of the likelihood function, we also update our guesses ω_0 and θ_0 by plugging in the maximum likelihood estimate and re-iterate the maximization of equation (2.5) 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 $\mathrm{d}(x)$; (ii) to test for the presence of residual spatial variation in $R(x)$ unexplained by the covariates $\mathrm{d}(x)$; 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 equation (2.2) 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 $\theta = (\alpha_1, \ldots, \alpha_J, \sigma_1, \ldots, \sigma_J)$ and $\omega = (\phi)$ is now given by the product of N one-dimensional integrals:

$$L(\omega, \theta) = [Y; \omega, \theta] = \prod_{i=1}^{N} \int_{\mathbb{R}} [U_i; \theta][Y_i|U_i; \theta]\, \mathrm{d}U_i. \tag{2.6}$$

To maximize the likelihood, each of the factors in the aforementioned product can then be approximated using numerical quadrature; we use a quasi Monte Carlo method, whereby the integrals in equation (2.2) are drawn deterministically based on the Halton sequence of support points [56].

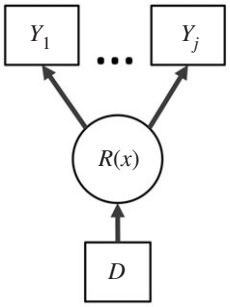

Figure 1. Directed acyclic graph of the geostatistical model of §2.2. $R(x)$ is the value of a spatially continuous stochastic process at location x. The outcome variables $Y_j : j = 1, \ldots, J$ are a set of indices that provide information about $R(x)$. The term D represents a set of explanatory variables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

After fitting the model in equation (2.6), we then estimate U_i using its predictive expectation:

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

To compute the $\hat{U}_i$, we plug in the maximum likelihood estimates for θ and ω. To explore the functional form of the relationship between $R(x_i)$ and the explanatory variables $\mathrm{d}(x_i)$, we plot the $\hat{U}_i$ against the values of $\mathrm{d}(x_i)$. On the basis of the empirical relationship observed in the scatterplots, we then introduce $\mathrm{d}(x_i)$ into the linear predictor for $R(x_i)$, leading to a first extension of the simplified model to

$$R(x_i) = \beta^{\top} \mathrm{d}(x_i) + U_i. \tag{2.7}$$

We can now use the model in equation (2.7) to pursue the second objective, i.e. testing for residual spatial correlation. We then re-fit the likelihood (equation (2.6)), without setting β 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 datasets, we compute the empirical variogram based on the $\hat{U}_i$ and use the resulting 10 000 variograms to compute 95% confidence intervals (CIs) 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 §2.2.

Finally, for the third objective, we test the null hypotheses $H_0 : \sigma_j = 0$, for $j = 1, \ldots, J$ using likelihood ratio tests. Note that, in this case, the conditions that give an asymptotic distribution of the likelihood ratio test based on a χ^2 distribution are not met because the value 0 is on the boundary of the parametric space of σ_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 $T(x) = \mathrm{d}(x)^{\top}\beta + \sqrt{\psi}S(x)$ at a set of prediction locations, $X^* = \{x_1^*, \ldots, x_H^*\}$. To predict

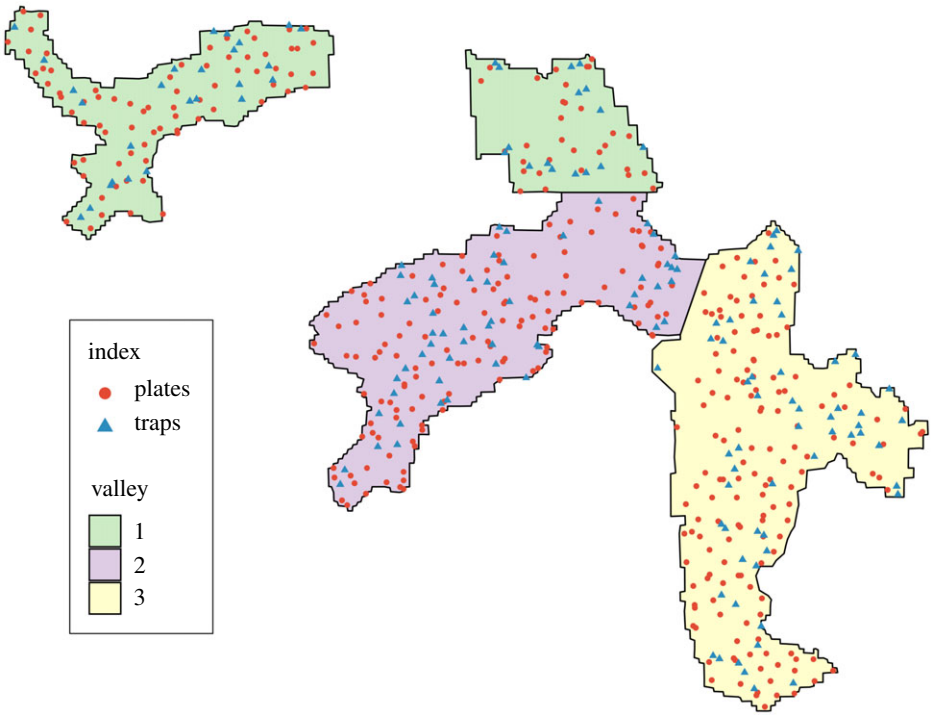

Figure 2. Map of the three valleys within the study site, Pau da Lima, with sampled locations for track plates and traps shown. Surveys for signs of rat infestation were conducted at all locations.

$T^* = (T(x_1^*), \ldots, T(x_H^*))$, we sample from its predictive distribution $[T^* \mid Y]$ as follows. We first simulate from $[R \mid Y; \theta, \omega]$ using the same sampling algorithm as for maximizing the likelihood in §2.2, with the parameters θ and ω fixed at their maximum likelihood estimates. After obtaining samples $r_{(b)}$, $b = 1, \ldots, B$, we then simulate from $[T^* \mid r_{(b)}]$, 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 $h = 1, \ldots, H$ and $b = 1, \ldots, B$, which are now samples drawn from $[T^* \mid Y]$. These can be used to compute any desired summary of the predictive distribution $[T^* \mid Y]$, such that the expectation

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

and the standard deviation

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

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 the excluding data for the jth 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 1000 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 15 years [30,32,41]. The study site at Pau da Lima is characterized 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 four consecutive 24 h trapping periods at each point (see Panti-May *et al.* [41]). After each 24 h period, trapping success and closure of the trap without a rat caught inside were recorded. Track plates were used for two consecutive 24 h periods at 415 locations [59] following the standardized protocol for placement and a survey developed and validated previously [6]. Five plates placed at each location in the shape of a 'five' on a die with 1 m spacing between each plate. After each 24 h 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

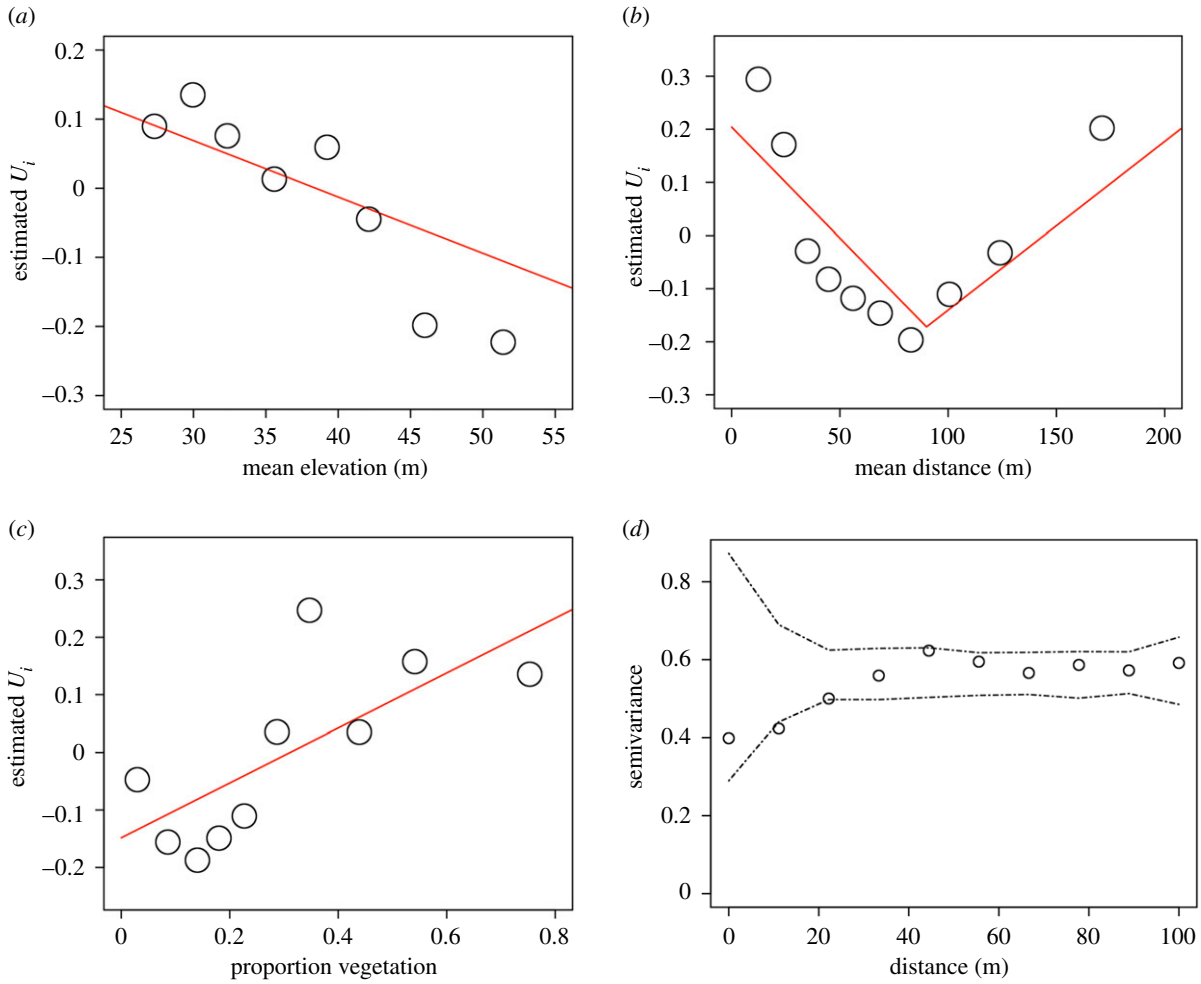

Figure 3. (*a*)–(*c*) Scatter-plots of the unstructured random variation, $\hat{U}_i$, against the covariates considered in the analysis. $\hat{U}_i$ are estimated using a non-spatial model, which excludes all covariates. The relationship between each covariate and $\hat{U}_i$ is shown with a red line and was estimated using univariable linear regression (with a linear piecewise spline knot included at 90 m for the distance to refuse covariate). (*a*) Elevation, (*b*) minimum distance to refuse dumps, and (*c*) land cover. (*d*) Variogram computed from $\hat{U}_i$ using a non-spatial model that includes all of the covariates; the dashed lines correspond to 95% CIs under the assumption of spatial independence.

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 m radius. Land cover data were created by the classification of Digital Globe's WorldView-2 satellite imagery (eight bands with resolution 0.5 m by 0.5 m taken on 17 February 2013) [61] using a maximum likelihood supervised algorithm. This was validated with ground truthing data collected from 20 randomly selected sites of size 5 m by 5 m.

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 ($j = 3$).

Owing to theft, violence and heavy rainfall, 993 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 of 1272 (6.7%) trapping-days were lost, of which 458 (38.6%) were found closed and empty after a 24 h period. This is a common issue and is owing 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 that 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 $t = 1$.

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

term	variable
β_1	elevation (m)
β_2	refuse distance (0–90 m)
β_3	refuse distance (>90 m)
β_4	proportion of land cover vegetation (%)
β_5	valley 2 (relative to valley 1)
β_6	valley 3 (relative to valley 1)

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

4. Results

4.1. Exploratory analysis

Following the steps outlined in §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 3*a*, we observe a negative linear relationship between elevation and *rattiness*. In figure 3*b*, we notice that, on average, higher values of *rattiness* are observed for distances from dumps less than 90 m. Finally, figure 3*c* shows that the mean proportion of land cover classified as vegetation (within a 30 m radius of a given sampling point) is approximately linearly and positively associated with *rattiness*.

On the basis of 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

$$R(x_i) = \sum_{l=1}^{6} \beta_l d_l(x_i) + U_i, \tag{4.1}$$

where each of the terms β_l corresponds to covariate effects as defined in table 1, d_l are the explanatory variables and U_i are the unstructured random effects.

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 3*d* shows the variogram for the $\hat{U}_i$ obtained from the spatially uncorrelated model (equation (4.1)). Most of the points of the empirical variogram lie inside the 95% tolerance band, but the variogram point at around 10 m is a highly unlikely value under the assumption of spatial independence. Because 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.

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, ϕ, of about 13 m indicates that the data exhibit spatial correlation after controlling for the explanatory variables. The estimate for ψ 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 β_l are consistent with the scatter-plots of figure 3. Both valleys 2 and 3 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.

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 4*a*) 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 4*b*. 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.

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 presented 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 in both sign and magnitude. The estimates for the scale of the spatial correlation, ϕ, from the 'traps and plates' and 'signs and plates' models are close to that of the full model, but the 'signs and traps' model had a substantially larger estimate of about 46 m with a much wider CI. 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 ψ very close to 1. We therefore fixed $\psi = 1$ for the 'traps and plates' and 'signs and 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 shows the maps of the relative differences for *rattiness* and figure 6 shows $S(x)$. The spatial predictions from the 'traps and plates' and 'signs and plates' models were more similar to

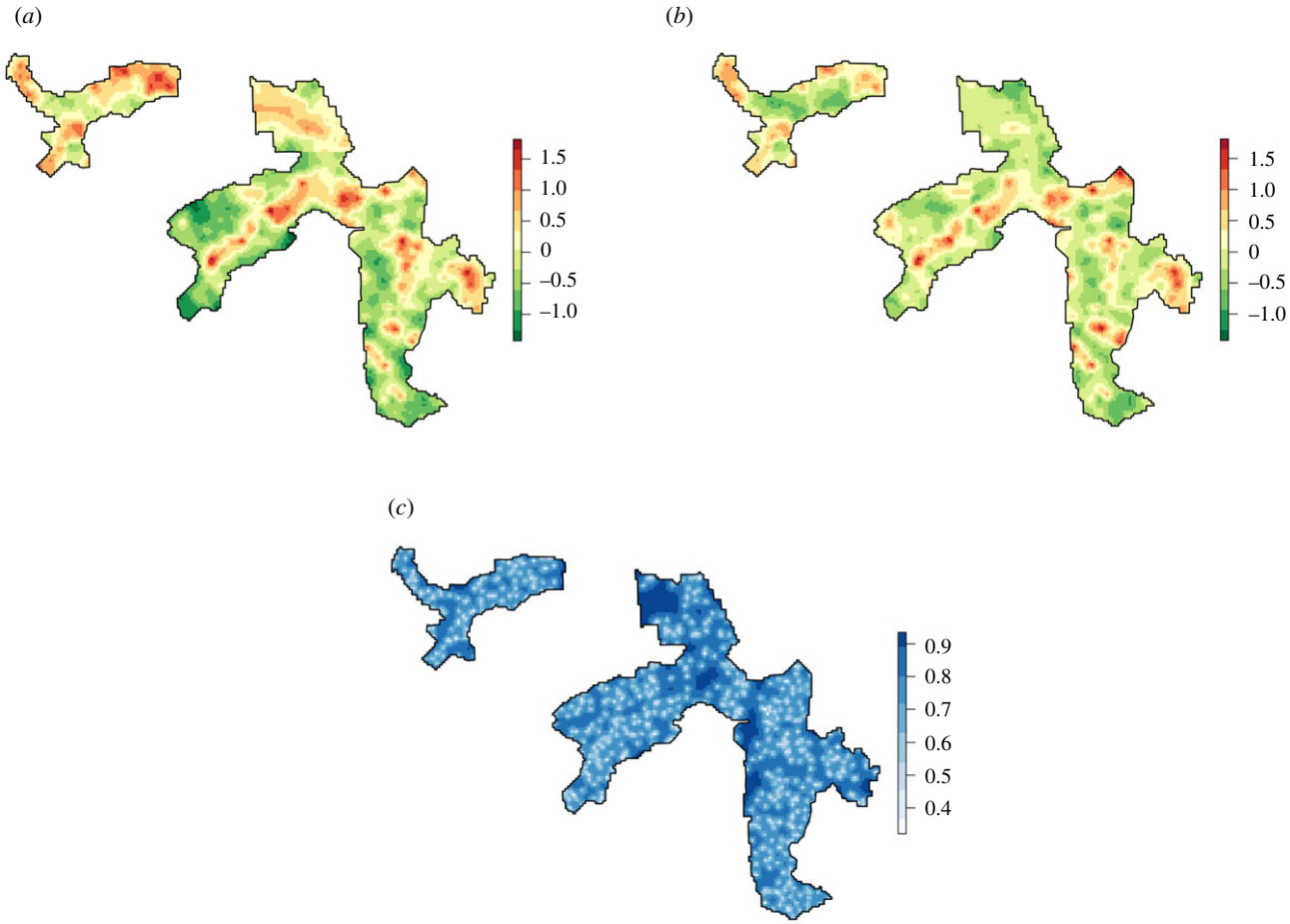

Figure 4. *Rattiness* model predictions (positive *rattiness* values indicate areas of high relative abundance). (*a*) Mean predicted values of *rattiness*, (*b*) mean predicted values of $S(x)$, and (*c*) standard deviation of predictions.

Table 2. Parameter estimates for the full model and the three two-indices models where α_1, α_2 and α_3 (and σ_1, σ_2 and σ_3) denote the coefficients for signs, traps and plates, respectively.

	estimate (95%CI)			
parameter	full model	signs and traps	traps and plates	signs and plates
α_1	−0.642 (−0.891,−0.425)	−0.508 (−0.997,−0.115)	—	−0.630 (−0.853,−0.419)
α_2	−2.684 (−3.078,−2.361)	−2.607 (−3.021,−2.099)	−2.456 (−2.849,−2.127)	—
α_3	−2.503 (−2.925,−2.144)	—	−2.567 (−3.066,−2.119)	−2.672 (−3.090,−2.305)
σ_1	0.747 (0.455, 1.045)	1.040 (0.615, 1.356)	—	0.641 (0.372, 0.906)
σ_2	0.920 (0.613, 1.183)	1.068 (0.546, 1.182)	0.863 (0.436, 1.133)	—
σ_3	1.896 (1.557, 2.179)	—	1.795 (1.348, 2.174)	1.877 (1.542, 2.163)
ϕ	13.432 (6.833, 21.172)	46.413 (7.692, 162.455)	11.306 (4.978, 17.325)	12.270 (6.920, 16.099)
ψ	0.878 (0.529, 1.000)	0.492 (0.161, 0.878)	—	—
β_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)
β_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)
β_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)
β_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)
β_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)
β_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)

those obtained from the full model as indicated by relative differences close to zero throughout the study area in both figures. By contrast, the predictions for the 'signs and 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.

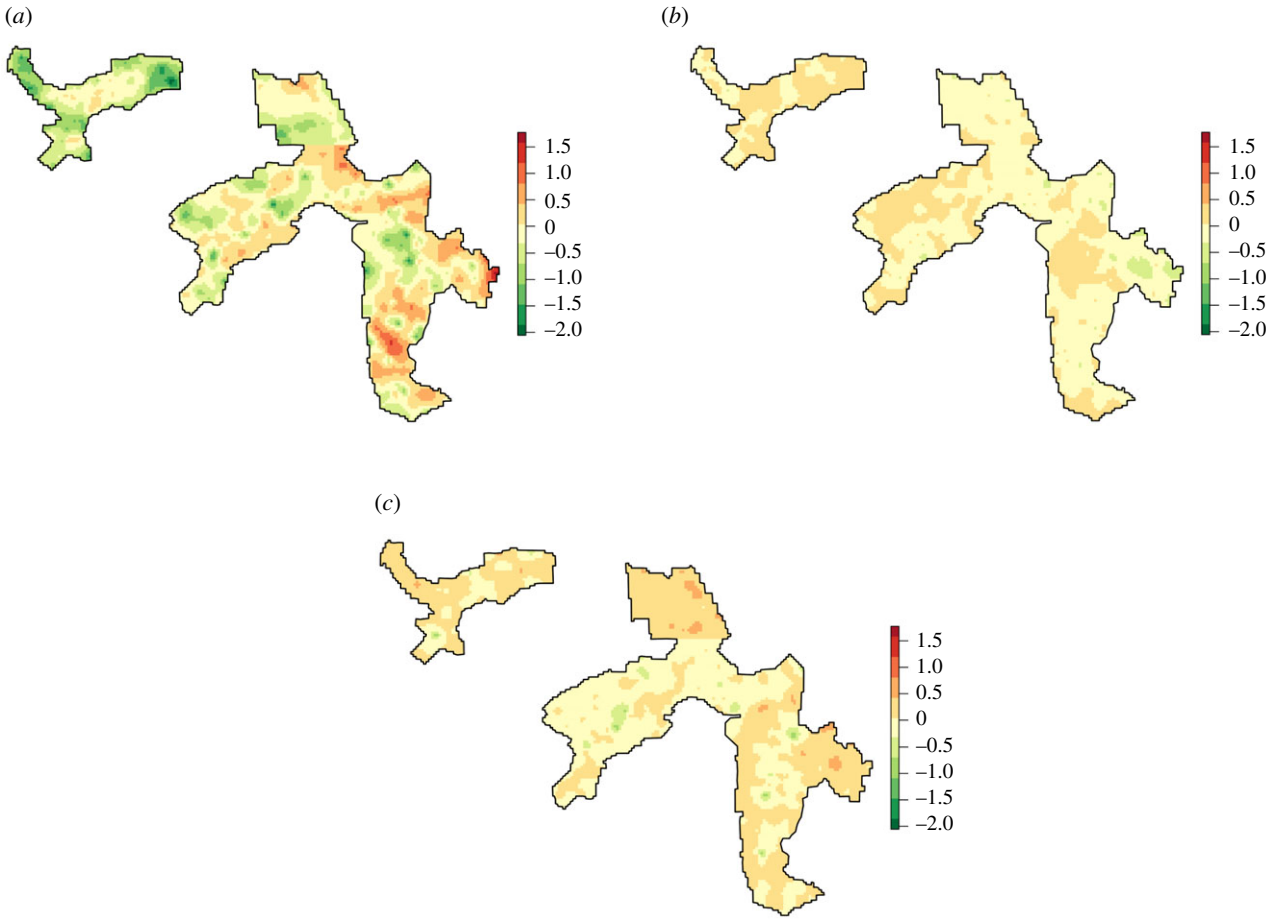

Figure 5. Model predictions for *rattiness* for each two-indices model relative to the full model (a value of zero indicates that there is no difference in the models' predicted values). (*a*) Signs and traps model, (*b*) traps and plates model, and (*c*) signs and plates model.

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 §2.4.)

model	SQ_M	SQ_{SD}
traps and plates	0.018	0.007
signs and plates	0.032	0.003
signs and traps	0.249	0.096

Table 3 presents the SQ_M and SQ_{SD} summaries used to quantify the changes in point predictions and standard deviations for *rattiness*; see §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*.

5. Discussion

In this article, 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 the 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 CIs for these estimates. A separate analysis that included elevation as the only covariate for *rattiness* (electronic supplementary material, S1) 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 CIs. 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

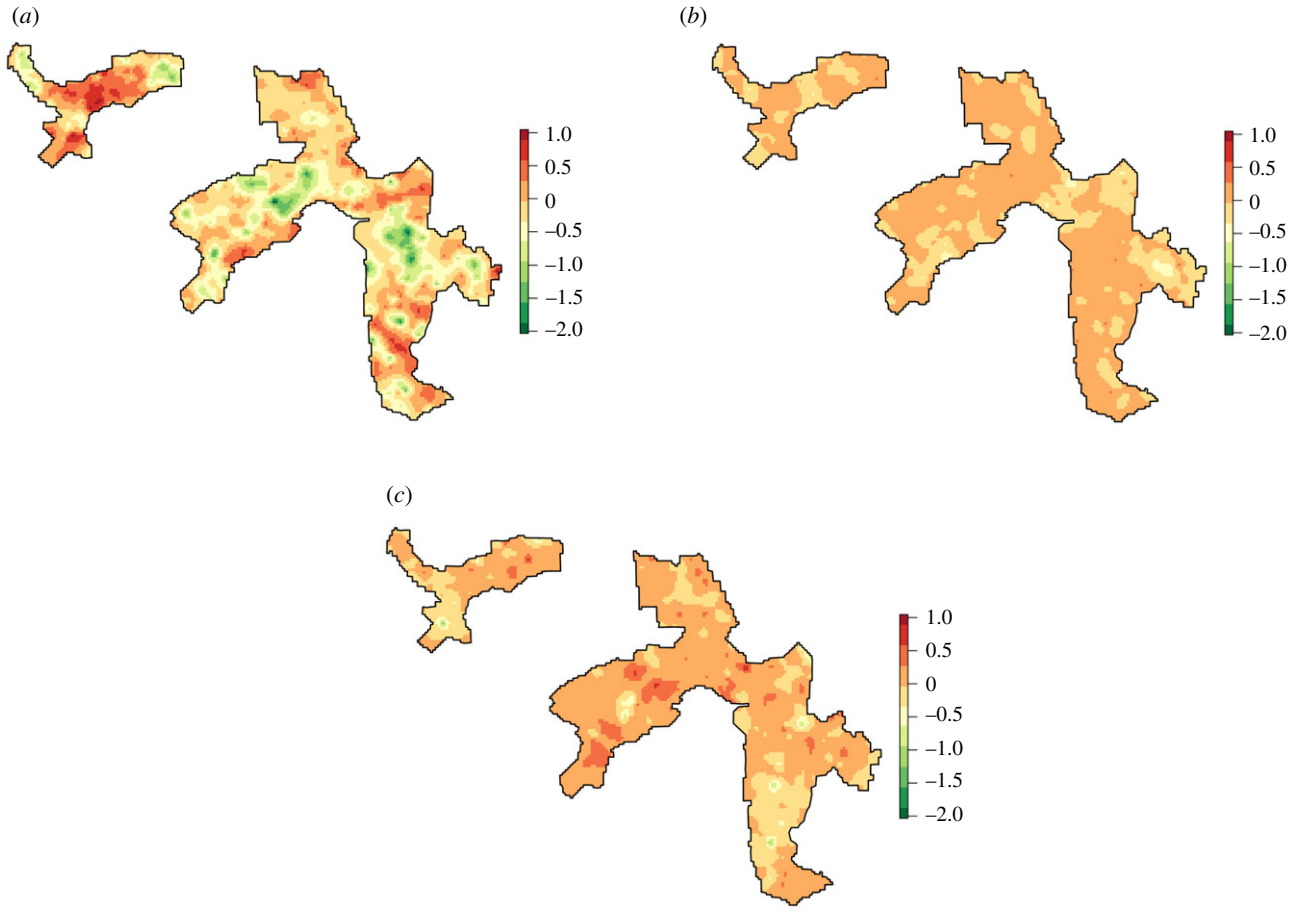

Figure 6. Model predictions for the spatial Gaussian process, $S(x)$, for each two-indices model relative to the full model (a value of zero indicates that there is no difference in the models' predicted $S(x)$ values). (*a*) Signs and traps model, (*b*) traps and plates model, and (*c*) signs and plates model.

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 owing 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, $\hat{\phi} \approx 13\,\text{m}$, corresponds to a spatial correlation range (the distance at which the correlation reduces to 0.05) of approximately 40 m. 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 25–150 m 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 because of 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 ψ 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. Owing 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 §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, $\text{R}(x)$, 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 both track plates and traps 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 the risk of human leptospirosis, the mechanism though 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 the 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 owing 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 are still widely collected for monitoring purposes without the requirement for these additional data sources.

Ethics. The ethics committee for the use of animals from the Oswaldo Cruz Foundation, Salvador, Brazil, approved the protocols used in this study (protocol number 003/2012), which adhered to the guidelines of the American Society of Mammalogists for the use of wild mammals in research [76] and the guidelines of the American Veterinary Medical Association for the euthanasia of animals [77]. These protocols were also approved by the Yale University's Institutional Animal Care and Use Committee (IACUC), New Haven, Connecticut (protocol number 2012–11498).

Data accessibility. R scripts for the *rattiness* model and all data needed for our analysis are available as part of this submission in the electronic supplementary material and also at https://github.com/maxeyre/Rattiness-1.

Authors' contributions. Conceived the project: A.I.K., E.G., F.C., M.B., M.G.R., M.T.E., P.J.D.; developed the model: E.G., M.T.E., P.J.D.; generation and preparation of case study data: F.C., F.N.S., H.K., J.P.T., K.P.H., M.T.E., S.S., T.S.A.C.P.; developed the statistical software and carried out the statistical analysis: M.T.E., E.G.; interpretation of results: E.G., M.T.E., P.J.D.; wrote the original draft of the manuscript: M.T.E.; comments on manuscript: all authors.

Competing interests. We declare we have no competing interest.

Funding. This work was supported by the Ministério da Ciência, Tecnologia e Inovação, the Conselho Nacional de Desenvolvimento Científico e Tecnológico; the Fundação de Amparo à Pesquisa do Estado da Bahia (grant no. 10206/20155); the Fundação Oswaldo Cruz; Research Councils UK, the Medical Research Council (grant nos. 1964635, MR/P024084/1 and MR/T029781/1); the Ministério da Saúde; the U.S. Department of Health and Human Services, the National Institutes of Health, the National Institute of Allergy and Infectious Diseases (grant nos. F31 AI114245, R01 AI052473, R01 TW009504, R25 TW009338 and U01 AI088752); the Secretaria Municipal de Saúde de Salvador-BA; and the Wellcome Trust (grant nos. 102330/Z/13/Z and 218987/Z/19/Z). MTE is in receipt of a studentship from the Medical Research Council, United Kingdom.

Reference

1. Kitron U. 1998 Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *J. Med. Entomol.* **35**, 435–445. (doi:10.1093/jmedent/35.4.435)
2. Cavia R, Rubn G, Virginia O. 2012 Techniques to estimate abundance and monitoring rodent pests in urban environments. In *Integrated pest management and pest control* (eds ML Larramendy, S Soloneski), pp. 147–172. Rijeka, Croatia: InTech.
3. Stephens PA, Pettorelli N, Barlow J, Whittingham MJ, Cadotte MW. 2015 Management by proxy? The use of indices in applied ecology. *J. Appl. Ecol.* **52**, 1–6. (doi:10.1111/1365-2664.12383)
4. Falcy MR, McCormick JL, Miller SA. 2016 Proxies in practice: calibration and validation of multiple indices of animal abundance. *J. Fish Wildlife Manage.* **7**, 117–128. (doi:10.3996/092015-JFWM-090)
5. O'Connell AF, Nichols JD, Karanth KU. 2010 *Camera traps in animal ecology: methods and analyses*. Berlin, Germany: Springer Science & Business Media.
6. Hacker KP, Minter A, Begon M, Diggle PJ, Serrano S, Reis MG, Childs JE, Ko AI, Costa F. 2016 A comparative assessment of track plates to quantify fine scale variations in the relative abundance of Norway rats in urban slums. *Urban Ecosyst.* **19**, 561–575. (doi:10.1007/s11252-015-0519-8)
7. Silver JB. 2007 *Mosquito ecology: field sampling methods*. Berlin, Germany: Springer Science & Business Media.
8. 2018 Malaria surveillance, monitoring & evaluation: a reference manual. See http://www.who.int/malaria/publications/atoz/9789241565578/en/.
9. Brown HE, Paladini M, Cook RA, Kline D, Barnard D, Fish D. 2008 Effectiveness of mosquito traps in measuring species abundance and composition. *J. Med. Entomol.* **45**, 517–521. (doi:10.1603/0022-2585(2008)45[517:EOMTIM]2.0.CO;2)
10. Boyer S, Foray C, Dehecq JS. 2014 Spatial and temporal heterogeneities of *Aedes albopictus* density in La Reunion Island: rise and weakness of entomological indices. *PLoS ONE* **9**, 1–12.
11. Tangena JAA, Thammavong P, Hiscox A, Lindsay SW, Brey PT. 2015 The human-baited double net trap: an alternative to human landing catches for collecting outdoor biting mosquitoes in Lao PDR. *PLoS ONE* **10**, 1–13.
12. Sinka ME, Golding N, Massey NC, Wiebe A, Huang Z, Hay SI, Moyes CL. 2016 Modelling the relative abundance of the primary African vectors of malaria before and after the implementation of indoor, insecticide-based vector control. *Malar. J.* **15**, 1–10. (doi:10.1186/s12936-016-1187-8)
13. Kraemer MUG *et al.* 2015 The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *eLife* **4**, e08347.
14. Dumonteil E, Gourbière S. 2004 Predicting Triatoma dimidiata abundance and infection rate: a risk map for natural transmission of Chagas disease in the Yucatán Peninsula of Mexico. *Am. J. Trop. Med. Hyg.* **70**, 514–519. (doi:10.4269/ajtmh.2004.70.514)
15. Stanton MC, Esterhuizen J, Tirados I, Betts H, Torr SJ. 2018 The development of high resolution maps of tsetse abundance to guide interventions against human African trypanosomiasis in northern Uganda. *Parasites and Vectors* **11**, 1–12. (doi:10.1186/s13071-018-2922-5)
16. Scholte RGC, Carvalho OS, Malone JB, Utzinger J, Vounatsou P. 2012 Spatial distribution of *Biomphalaria* spp., the intermediate host snails of

Schistosoma mansoni, in Brazil. *Geospatial Health* **6**(Suppl. 3), 95–101. (doi:10.4081/gh.2012.127)

17. Ferro C, López M, Fuya P, Lugo L, Cordovez JM, González C. 2015 Spatial distribution of sand fly vectors and eco-epidemiology of cutaneous leishmaniasis transmission in Colombia. *PLoS ONE* **10**, 1–16. (doi:10.1371/journal.pone.0139391)
18. Winters AM, Bolling BG, Beaty BJ, Blair CD, Eisen RJ, Meyer AM, Pape WJ, Moore CG, Eisen L. 2008 Combining mosquito vector and human disease data for improved assessment of spatial west Nile virus disease risk. *Am. J. Tropic. Med. Hyg.* **78**, 654–665. (doi:10.4269/ajtmh.2008.78.654)
19. Himsworth CG, Bidulka J, Parsons KL, Feng AY, Tang P, Jardine CM, Kerr T, Mak S, Robinson J, Patrick DM. 2013 Ecology of *Leptospira interrogans* in Norway rats (*Rattus norvegicus*) in an inner-city neighborhood of Vancouver, Canada. *PLoS Negl. Trop. Dis.* **7**, e2270. (doi:10.1371/journal.pntd.0002270)
20. Wilschut LI *et al.* 2013 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. *Int. J. Appl. Earth Obs. Geoinf.* **23**, 81–94. (doi:10.1016/j.jag.2012.11.007)
21. Carbajo AE, Pardiñas UFJ. 2007 Spatial distribution model of a hantavirus reservoir, the long-tailed colilargo (*Oligoryzomys longicaudatus*), in Argentina. *J. Mammal.* **88**, 1555–1568. (doi:10.1644/06-MAMM-A-183R1.1)
22. Morand S *et al.* 2015 Assessing the distribution of disease-bearing rodents in human-modified tropical landscapes. *J. Appl. Ecol.* **52**, 784–794. (doi:10.1111/1365-2664.12414)
23. Diggle PJ, Tawn JA, Moyeed RA. 1998 Model-based geostatistics. *J. R. Stat. Soc.: Ser. C* (*Applied Statistics*) **47**, 299–350. (doi:10.1111/1467-9876.00113)
24. Diggle PJ, Giorgi E. 2019 *Model-based geostatistics for global public health: methods and applications*. Boca Raton, FL: CRC Press.
25. Royle JA, Kéry M, Gautier R, Schmid H. 2007 Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecol. Monogr.* **77**, 465–481. (doi:10.1890/06-0912.1)
26. Rivoirard J, Simmonds J, Foote KG, Fernandes P, Bez N. 2008 *Geostatistics for estimating fish abundance*. Hoboken, NJ: Wiley.
27. Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, Stein C, Abela-Ridder B, Ko AI. 2015 Global morbidity and mortality of leptospirosis: a systematic review. *PLoS Negl. Trop. Dis.* **9**, e0003898.
28. Torgerson PR *et al.* 2015 Global burden of leptospirosis: estimated in terms of disability adjusted life years. *PLoS Negl. Trop. Dis.* **9**, 1–14. (doi:10.1371/journal.pntd.0004122)
29. Haake DA, Levett PN. 2015 Leptospirosis in humans. In *Leptospira and leptospirosis* (ed. B Adler), pp. 65–97. Berlin, Germany: Springer.
30. Costa F, Wunder EA Jr, De Oliveira D, Bisht V, Rodrigues G, Reis MG, Ko AI, Begon M, Childs JE. 2015 Patterns in *Leptospira* shedding in Norway rats (*Rattus norvegicus*) from Brazilian slum communities at high risk of disease transmission. *PLoS Negl. Trop. Dis.* **9**, 1–14.
31. Reis RB *et al.* 2008 Impact of environment and social gradient on *Leptospira* infection in urban slums. *PLoS Negl. Trop. Dis.* **2**, 11–18.
32. Hagan JE *et al.* 2016 Spatiotemporal determinants of urban leptospirosis transmission: four-year prospective cohort study of slum residents in Brazil. *PLoS Negl. Trop. Dis.* **10**, 1–17. (doi:10.1371/journal.pntd.0004275)
33. Mwachui MA, Crump L, Hartskeerl R, Zinsstag J, Hattendorf J. 2015 Environmental and behavioural determinants of leptospirosis transmission: a systematic review. *PLoS Negl. Trop. Dis.* **9**, 1–15. (doi:10.1371/journal.pntd.0003843)
34. Costa F *et al.* 2014 Influence of household rat infestation on *Leptospira* transmission in the urban slum environment. *PLoS Negl. Trop. Dis.* **8**, e3338. (doi:10.1371/journal.pntd.0003338)
35. Sarkar U *et al.* 2002 Population-based case-control investigation of risk factors for leptospirosis during an urban epidemic. *Am. J. Trop. Med. Hyg.* **66**, 605–10. (doi:10.4269/ajtmh.2002.66.605)
36. Vanasco NB, Sequeira MD, Sequeira G, Tarabla HD. 2003 Associations between *Leptospira*l infection and seropositivity in rodents and environmental characteristics in Argentina. *Prev. Vet. Med.* **60**, 227–235. (doi:10.1016/S0167-5877(03)00144-2)
37. Santos NdJ, Sousa E, Reis MG, Ko AI, Costa F. 2017 Rat infestation associated with environmental deficiencies in an urban slum community with high risk of leptospirosis transmission. *Cadernos de Saúde Pública* **33**, 1–13. See www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2017000205003&lng=en&tlng=en..
38. Koizumi N *et al.* 2009 Human leptospirosis cases and the prevalence of rats harbouring *Leptospira interrogans* in urban areas of Tokyo, Japan. *J. Med. Microbiol.* **58**, 1227–1230. (doi:10.1099/jmm.0.011528-0)
39. Maciel EA, de Carvalho AL, Nascimento SF, de Matos RB, Gouveia EL, Reis MG, Ko AI. 2008 Household transmission of *Leptospira* infection in urban slum communities. *PLoS Negl. Trop. Dis.* **2**, e154. (doi:10.1371/journal.pntd.0000154)
40. Minter A, Diggle PJ, Costa F, Childs J, Ko AI, Begon M. 2018 A model for *Leptospira* dynamics and control in the Norway rat (*Rattus norvegicus*) the reservoir host in urban slum environments. *Epidemics* **25**, 26–34. (doi:10.1016/j.epidem.2018.05.002)
41. Panti-May JA *et al.* 2016 A two-year ecological study of norway rats (*Rattus norvegicus*) in a Brazilian urban slum. *PLoS ONE* **11**, e0152511.
42. Fletcher RJ, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM. 2019 A practical guide for combining data to model species distributions. *Ecology* **100**, 1–15.
43. Dobbie MJ, Dail D. 2013 Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecol. Indic.* **29**, 270–277. (doi:10.1016/j.ecolind.2012.12.025)
44. Engeman RM. 2005 Indexing principles and a widely applicable paradigm for indexing animal populations. *Wildlife Res.* **32**, 203–210. (doi:10.1071/WR03120)
45. Buckland ST, Studeny AC, Magurran AE, Illian JB, Newson SE. 2011 The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere* **2**, art100. (doi:10.1890/ES11-00186.1)
46. Heink U, Kowarik I. 2010 What are indicators? On the definition of indicators in ecology and environmental planning. *Ecol. Indic.* **10**, 584–593. (doi:10.1016/j.ecolind.2009.09.009)
47. Reza MIH. 2014 Importance and considerations for the development of a composite index of ecological integrity for ecological management. *Int. J. Ecol. Dev.* **28**, 32–48.
48. Vačkář D, Ten Brink B, Loh J, Baillie JEM, Reyers B. 2012 Review of multispecies indices for monitoring human impacts on biodiversity. *Ecol. Indic.* **17**, 58–67. (doi:10.1016/j.ecolind.2011.04.024)
49. Dey S, Delampady M, Parameshwaran R, Kumar NS, Srivathsa A, Karanth KU. 2017 Bayesian methods for estimating animal abundance at large spatial scales using data from multiple sources. *J. Agric. Biol. Environ. Stat.* **22**, 111–139. (doi:10.1007/s13253-017-0276-7)
50. Amiri K, Shabanipour N, Eagderi S. 2017 Using kriging and co-kriging to predict distributional areas of Kilka species (*Clupeonella* spp.) in the southern Caspian Sea. *Int. J. Aquat. Biol.* **5**, 108–113.
51. Wang Y, Naumann U, Wright ST, Warton DI. 2012 Mvabund: an R package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.* **3**, 471–474. (doi:10.1111/j.2041-210X.2012.00190.x)
52. Ovaskainen O, Soininen J. 2011 Making more out of sparse data: hierarchical modeling of species communities. *Ecology* **92**, 289–295. (doi:10.1890/10-1251.1)
53. Pacifici K, Reich BJ, Miller DAW, Pease BS. 2019 Resolving misaligned spatial data with integrated species distribution models. *Ecology* **100**, 1–15. (doi:10.1002/ecy.2747)
54. Christensen OF. 2004 Monte Carlo maximum likelihood in model-based geostatistics. *J. Comput. Graph. Stat.* **13**, 702–718. (doi:10.1198/106186004X2525)
55. Giorgi E, Diggle PJ. 2017 PrevMap: an R package for prevalence mapping. *J. Stat. Softw.* **78**, 1–29. (doi:10.18637/jss.v078.i08)
56. Caflisch RE. 1998 Monte Carlo and quasi-Monte Carlo methods. *Acta Numer.* **7**, 1–49. (doi:10.1017/S0962492900002804)
57. Self SG, Liang KY. 1987 Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **82**, 605–610. (doi:10.1080/01621459.1987.10478472)
58. Hacker KP *et al.* 2020 Influence of rainfall on *Leptospira* infection and disease in a tropical urban setting, Brazil. *Emerg. Infect. Dis.* **26**, 311–314. (doi:10.3201/eid2602.190102)

59. Souza FN *et al.* In preparation. Predicting the variation in the temporal and spatial presence of urban rats in tropical slums.
60. Centers for Disease Control and Prevention. 2006 *Integrated pest management: conducting urban rodent surveys*. Atlanta, GA: US Department of Health and Human Services.
61. Hacker KP *et al.* 2016 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, p. 2020. Atlanta, GA, USA.
62. De Masi E, Vilaça P, Razzolini MTP. 2009 Environmental conditions and rodent infestation in Campo Limpo district, São Paulo municipality, Brazil. *Int. J. Environ. Health Res.* **19**, 1–16. (doi:10.1080/09603120802126670)
63. Engeman RM, Witmer GW. 2000 IPM strategies: indexing difficult to monitor populations of pest species. In *Proceedings of the 19th Vertebrate Pest Conference,* 6–9 March 2000, San Diego, CA, (eds TP Salmon, AC Crabb). Davis, CA: University of California Davis.
64. Traweger D, Travnitzky R, Moser C, Walzer C, Bernatzky G. 2006 Habitat preferences and distribution of the brown rat (*Rattus norvegicus* Berk.) in the city of Salzburg (Austria): implications for an urban rat management. *J. Pest Sci.* **79**, 113–125. (doi:10.1007/s10340-006-0123-z)
65. Feng AYT, Himsworth CG. 2014 The secret life of the city rat: a review of the ecology of urban Norway and black rats (*Rattus norvegicus* and *Rattus rattus*). *Urban Ecosyst.* **17**, 149–162. (doi:10.1007/s11252-013-0305-4)
66. Gardner-Santana LC, Norris DE, Fornadel CM, Hinson ER, Klein SL, Glass GE. 2009 Commensal ecology, urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (*Rattus norvegicus*). *Mol. Ecol.* **18**, 2766–2778. (doi:10.1111/j.1365-294X.2009.04232.x)
67. Himsworth CG, Jardine CM, Parsons KL, Feng AYT, Patrick DM. 2014 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* **9**, e91654.
68. Hoverman JT, Searle CL. 2016 Behavioural influences on disease risk: implications for conservation and management. *Anim. Behav.* **120**, 263–271. (doi:10.1016/j.anbehav.2016.05.013)
69. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, Lloyd-Smith JO. 2017 Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510. (doi:10.1038/nrmicro.2017.45)
70. Hopkins HL, Kennedy ML. 2004 An assessment of indices of relative and absolute abundance for monitoring populations of small mammals. *Wildlife Soc. Bull.* **32**, 1289–1296. (doi:10.2193/0091-7648(2004)032[1289:AAOIOR]2.0.CO;2)
71. Baldwin RA, Quinn N, Davis DH, Engeman RM. 2014 Effectiveness of rodenticides for managing invasive roof rats and native deer mice in orchards. *Environ. Sci. Pollut. Res.* **21**, 5795–5802. (doi:10.1007/s11356-014-2525-4)
72. Towns DR, Atkinson IAE, Daugherty CH. 2006 Have the harmful effects of introduced rats on islands been exaggerated? *Biol. Invasions* **8**, 863–891. (doi:10.1007/s10530-005-0421-z)
73. Ruffell J, Innes J, Didham RK. 2015 Efficacy of chew-track-card indices of rat and possum abundance across widely varying pest densities. *N. Z. J. Ecol.* **39**, 87–92.
74. Nkosi SE, Adam E, Barrett AS, Brown LR. 2019 A synopsis of field and remote sensing based methods for studying African elephant (*Loxodonta africana*) impact on woody vegetation in Africa. *Appl. Ecol. Environ. Res.* **17**, 4045–4066. (doi:10.15666/aeer/1702_40454066)
75. Sollmann R, Mohamed A, Samejima H, Wilting A. 2013 Risky business or simple solution: relative abundance indices from camera-trapping. *Biol. Conserv.* **159**, 405–412. (doi:10.1016/j.biocon.2012.12.025)
76. Sikes RS, Gannon WL, Animal Care and Use Committee of the American Society of Mammalogists. 2011 Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *J. Mammal.* **92,** 235–253.
77. Leary S, Underwood W, Lilly E, Anthony R, Cartner S, Corey D, et al. 2013 AVMA guidelines for the euthanasia of animals. Schaumburg, IL: American Veterinary Medical Association.