Ecological Applications, 0(0), 2021, e02329

Spatially clustered count data provide more efficient search strategies in invasion biology and disease control

Michael C. A. Stevens [1,2,7], Sally C. Faulkner [1], André B. B. Wilke [3], John C. Beier [3], Chalmers Vasquez [4], William D. Petrie [4], Hannah Fry [2], Richard A. Nichols [1], Robert Verity [5], and Steven C. Le Comber [1,6]

[1]*School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS UK*
[2]*Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ UK*
[3]*Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, Florida 33136 USA*
[4]*Miami-Dade County Mosquito Control Division, Miami, Florida 33178 USA*
[5]*Department of Infectious Disease Epidemiology, MRC Centre for Global Infectious Disease Analysis, Imperial College London, London W2 1PG UK*

Citation: Stevens, M. C. A., S. C. Faulkner, A. B. B. Wilke, J. C. Beier, C. Vasquez, W. D. Petrie, H. Fry, R. A. Nichols, R. Verity, and S. C. Le Comber. 2021. Spatially clustered count data provide more efficient search strategies in invasion biology and disease control. Ecological Applications 00(00):e02329. 10.1002/eap.2329

Abstract. Geographic profiling, a mathematical model originally developed in criminology, is increasingly being used in ecology and epidemiology. Geographic profiling boasts a wide range of applications, such as finding source populations of invasive species or breeding sites of vectors of infectious disease. The model provides a cost-effective approach for prioritizing search strategies for source locations and does so via simple data in the form of the positions of each observation, such as individual sightings of invasive species or cases of a disease. In doing so, however, classic geographic profiling approaches fail to make the distinction between those areas containing observed absences and those areas where no data were recorded. Absence data are generated via spatial sampling protocols but are often discarded during the inference process. Here we construct a geographic profiling model that resolves these issues by making inferences via count data, analyzing a set of discrete sentinel locations at which the number of encounters has been recorded. Crucially, in our model this number can be zero. We verify the ability of this new model to estimate source locations and other parameters of practical interest via a Bayesian power analysis. We also measure model performance via real-world data in which the model infers breeding locations of mosquitoes in bromeliads in Miami-Dade County, Florida, USA. In both cases, our novel model produces more efficient search strategies by shifting focus from those areas containing observed absences to those with no data, an improvement over existing models that treat these areas equally. Our model makes important improvements upon classic geographic profiling methods, which will significantly enhance real-world efforts to develop conservation management plans and targeted interventions.

Key words: Bayesian parameter estimation; Dirichlet process; epidemiology; finite mixture model; mapping; mosquito.

Introduction

Geographic profiling is a tool originally used in criminology in cases of serial crime such as murder, rape, or arson, to find the most likely area(s) for the offender's anchor point(s) (usually a home, but sometimes a workplace or relative's home), using as input the locations of crimes associated with that offender (Rossmo 2000). It is designed to deal with cases of information overload, where there are insufficient resources to deal with the large numbers of suspects typical in investigations of serial crime (for example, the Yorkshire Ripper enquiry in the UK generated 268,000 names and 5.4 million vehicle registrations [Doney 1990]).

In criminology, geographic profiling uses the spatial locations associated with the crimes (e.g., victim encounter sites, body dump sites, weapon dump sites) to produce a three-dimensional probability surface that can be overlaid on a map of the study area to produce a geographic profile. Suspects are prioritized according to the height of their anchor point(s) on the surface (Rossmo 2000). Geographic profiling is widely used by law enforcement agencies around the world (Rossmo 2012), but more recently has been applied to cases in ecology and epidemiology where spatial locations are associated with sightings of an invasive species or an instance of an infectious disease (Table 1). Geographic profiling boasts a variety of successful applications from invasion biology

Manuscript received 20 August 2020; revised 23 October 2020; accepted 6 December 2020. Corresponding Editor: Trenton W. J. Garner.
[6]Deceased.
[7] E-mail: m.stevens@qmul.ac.uk

Table 1. Terminology used in geographic profiling and species distribution models alongside joint terms adopted in this study.

Discipline and examples	Event	Encounter	No encounter	Source location	Sentinel site
Ecology					
Faulkner et al. (2016)	Invasive species	Capture	Empty trap	Nesting location	Trap
Chandler and Royle (2013)	Animal	Observed individual	Nothing observed	Activity center	Trap: single, multi-level, proximity
Epidemiology					
Verity et al. (2014)	Disease host	Positive test result	Negative test result	Source of outbreak	Patient postcode
Criminology					
Rossmo et al. (2014)	Criminal	Crime	No crime	Anchor point	Potential crime site

(Stevenson et al. 2012, Papini et al. 2013, Faulkner et al. 2016, Cerri et al. 2020, Heald et al. 2019) to animal behavior (Le Comber et al. 2006, Martin et al. 2009, Raine et al. 2009, Faulkner et al. 2015), human–wildlife conflict (Faulkner et al. 2018, Struebig et al. 2018), and epidemiology (Le Comber et al. 2011, Verity et al. 2014, Smith et al. 2015).

There are a number of geographic profiling models, from the Criminal Geographic Targeting (CGT) algorithm used in criminology (Rossmo 1993, Rossmo et al. 2014, Butkovic et al. 2018) to explicitly Bayesian models (O'Leary 2009, 2010, Mohler and Short 2012) and, more recently, the Dirichlet Process Mixture (DPM) model (Verity et al. 2014, Faulkner et al. 2016). However, all these models have one thing in common in that they use point-pattern data only: a finite collection of longitudinal/latitudinal points each associated with a single instance of crime or sighting of an invasive species etc.

By considering count data, we can make an important distinction between evidence of absence and absence of evidence. In an ecological context, this might relate to areas where traps were set but failed to catch any animals and areas where no traps were set; in criminology, between areas where crimes could have been committed but were not and areas where no information was recorded (such as outside a jurisdictional boundary); and in epidemiology, between areas where people were tested and found negative, and areas where no one was tested.

There are existing models in ecology that can use count data to infer parameters of biological interest. For example, spatially explicit capture recapture models aim to estimate the underlying population density in a study area given the locations of discrete traps with associated counts (Borchers and Efford 2008, Chandler and Royle 2013). These models even go so far as to estimate an individual's "activity center" a latent variable synonymous to "source location" or "anchor point" used throughout geographic profiling literature. These models, however, assume each individual from a species is associated with its own unique activity center of which are estimated from the data. The DPM model however, does not assume this and is built to deal with the complex problem of partitioning individuals into spatial clusters of which each cluster is governed by a single "source location" (Verity et al. 2014).

In ecology, it is often common for count data to exhibit over-dispersion, that is, data stray from the assumed equal mean and variance, a standard to those modeling count data via some underlying expectation for a Poisson density. This over-dispersion can be caused by a range of factors such as sampling, aggregation, environmental variability or a combination of the above (Lindén and Mäntyniemi 2011). As an alternative, count data can be modeled such that variance in counts is a linear or quadratic function of the mean (Ver Hoef and Boveng 2007). Hence some consideration is needed for over-dispersion when building a geographic profiling model that makes inferences via count data.

In this study, we address the gap in existing geographic profiling models by developing a fully Bayesian geographic profiling model for analyzing count data. We do this by calculating the likelihood of a particular number of crimes (or captures, or positive tests) at a given location, which can include zero. In addition to including count data in the model's likelihood, we will demonstrate how this leads to, for the first time, an estimation of the expected population size over a search area and time period. This is a parameter of consistent interest spanning disciplines, from criminology, in estimating the number of prostitutes or migrating fugitives (Rossmo and Routledge 1990), to ecology, in estimating the population size of many avian species (Royle 2004).

The performance of the new model is tested first by a Bayesian analogue of a power analysis of simulated data. We then demonstrate how this model can be expanded to deal with over-dispersed count data, and test such a model on a real-world data set in which we infer breeding site locations of the mosquito *Aedes aegypti*, one of the primary transmitters of Zika virus across the globe (Hayes 2009, Hennessey et al. 2016). We investigate model behavior when each search for bromeliad source locations given (1) the DPM model using repeat point-pattern data of traps yielding mosquitoes and (2) the negative binomial model using the full count data, including those with no encounters. The model excelled when making inferences based on simulated and real-world data; search strategies based on count data shifted attention from those areas containing zeros, to those containing no information.

Methods

A Poisson geographic profiling model

The Poisson model begins by assuming K sources, with locations $\boldsymbol{\mu}_k = (\mu_x, \mu_y)$ for k in 1:K drawn from some suitable prior distribution, F. Here we follow (O'Leary 2010) in assuming that F is defined over a two-dimensional grid of cells, allowing the prior probability mass to be defined separately for each cell (for example, we often want zero probability over water bodies). Next, we assume there is some expectation, λt, on the number of events, both encountered and unencountered, in the study area, where λ is the expected number of events over the search area per unit time and t is the time interval with which data were collected. From this expectation we make a Poisson draw to obtain the total number of events, N, in the study area. Explicitly, an event is the existence of an invasive species, a host of a disease or a criminal in our search area.

Each event originates from a single source with equal probability $1/K$, and the source from which event i originates can be written as c_i in 1:K. The spatial location of event i, denoted $\mathbf{x}_i$, is drawn from a dispersal distribution centered on its source. Here we assume a bivariate Normal distribution with mean $\boldsymbol{\mu}_{c_i}$ and variance $\sigma^2_{c_i}$, and zero correlation between dimensions. This is consistent with previous geographic profiling studies that recognize the probability of encountering an event is defined over two-dimensional space as opposed to spatial capture recapture models that consider a univariate half-normal distribution between source and event (Efford 2004).

Unlike the DPM model, we do not assume that every event is encountered. Instead we assume that there are m sentinel sites, denoted $\mathbf{s}_j$ for j in 1:m, within the study area and that events are only encountered if they fall within a distance ρ from one of these sites. A sentinel site could take on many forms as shown in Table 1, biologically, these could refer to camera traps, hair snares, or bioacoustics (Royle et al. 2018). In this study, sentinel sites can encounter any non-negative integer of events akin to multi-catch traps in ecology (Borchers 2012), leading us to our count data. We make the model fully Bayesian by placing suitable priors on the remaining unknown quantities of interest. The complete model can be written

Likelihood

$$c_i \sim \text{Categorical}(1/K), \quad \text{for } i = 1:N,$$

$$\mathbf{x}_i \sim \text{Normal}\left(\boldsymbol{\mu}_{c_i}, \boldsymbol{I}_2\sigma^2_{c_i}\right), \quad \text{for } i = 1:N,$$

$$n_j = \left\{\mathbf{x}_i | d_{\text{E}}\left(\mathbf{x}_i, \mathbf{s}_j\right) < \rho\right\}, \quad \text{for } j = 1:m,$$

Priors

$$\boldsymbol{\mu}_{c_i} \sim F, \quad \text{for } c_i = 1:K,$$

$$\sigma_{c_i} \sim \text{Log} - \text{Normal}\left(\gamma, \delta\right), \quad \text{for } c_i = 1:K,$$

$$N \sim \text{Poisson}\left(\lambda t\right)$$

$$\lambda \sim \text{Gamma}(\zeta, \eta), \tag{1}$$

where $\mathbf{I}_2$ is the two-dimensional identity matrix, and $d_{\text{E}}(\mathbf{x}_i, \mathbf{s}_j)$ the Euclidian distance between points $\mathbf{x}_i$ and $\mathbf{s}_j$. When performing inference, we only have access to the final counts n_j at each of the m sentinel sites, and not the raw data $\mathbf{x}_i$ for i in 1:N.

Now we need to calculate the probability of the observed n_j given the parameters $\{\boldsymbol{\mu}_{c_i}\}$, $\{\boldsymbol{\sigma}_{c_i}\}$ and λt (the likelihood). The probability that an event is observed is equal to the probability that it falls within a distance ρ of a sentinel site, which can be obtained by integrating the dispersal distribution over the ball $B_\rho(\mathbf{s}_j)$ of radius ρ centered on $\mathbf{s}_j$. In general, this integral will not have a simple analytical solution, but under certain conditions we can approximate the volume of integration by a cylinder centered on $\mathbf{s}_j$ with radius ρ and height equal to the dispersal distribution at the central point

$$\Pr(d_{\text{E}}(\mathbf{x}_i, \mathbf{s}_j) < \rho | \boldsymbol{\mu}, \sigma, c_i) = \iint\limits_{B_\rho(\mathbf{s}_j)} f_{\text{BN}}(x, y | \boldsymbol{\mu}_{c_i}, \boldsymbol{I}_2\sigma^2_{c_i}) dx dy,$$

$$\approx \pi\rho^2 f_{\text{BN}}\left(\boldsymbol{s}_j | \boldsymbol{\mu}_{c_i}, \boldsymbol{I}_2\sigma^2_{c_i}\right) \tag{2}$$

where f_{BN} $(\mathbf{s}_j|\boldsymbol{\mu},$ $\boldsymbol{I}_2\sigma^2_{c_i})$ is the density of the bivariate normal distribution at sentinel site j with mean $\boldsymbol{\mu}_{c_i}$ and covariance matrix $\boldsymbol{I}_2\sigma^2_{c_i}$. The validity of this approximation is explored in detail in Appendix S1. The total probability of being detected by sentinel site j can be obtained by averaging over all sources, leading to the following expression, which we define as θ_j for convenience

$$\begin{aligned}\theta_j &\equiv \Pr\left(d_{\text{E}}\left(\mathbf{x}_i, \mathbf{s}_j\right) < \rho | \left\{\boldsymbol{\mu}_{c_i}\right\}, \left\{\sigma_{c_i}\right\}\right) \\ &= \frac{\pi\rho^2}{K} \sum_{c_i=1}^{K} f_{\text{BN}}\left(\mathbf{s}_j | \boldsymbol{\mu}_{c_i}, \boldsymbol{I}_2\sigma^2_{c_i}\right).\end{aligned} \tag{3}$$

Given a Poisson prior with rate λt is applied to the total number of events N and that every event has the same independent probability of being detected given by Eq. 3, it follows that the probability of detecting n_j events at sentinel site j is independently Poisson distributed with rate $\lambda t\theta_j$. The likelihood is obtained by multiplying this Poisson probability over all sentinel sites

$$\Pr(\mathbf{n} | \lambda t, \boldsymbol{\theta}) = \prod_{j=1}^{m} \frac{\left(\lambda t\theta_j\right)^{n_j} e^{-\lambda t\theta_j}}{n_j!}. \tag{4}$$

Here we assume the unit of time is the interval in which the data were collected and thus set t equal to 1.

To account for potential over-dispersion in count data we can alter the likelihood in (4) as follows. We adopt a re-parametrized negative binomial density and introduce

a dispersion parameter α such that count n_j is drawn from this density with mean $\lambda t\theta_j$ and variance $\lambda t\theta_j + \alpha(\lambda t\theta_j)^2$ (Lindén and Mäntyniemi 2011). Under a negative binomial model, the likelihood in Eq. 4 switches to

$$\Pr(\mathbf{n}|\lambda t, \boldsymbol{\theta}, r) = \prod_{j=1}^{m} \frac{\Gamma(r+n_j)}{n_j!\Gamma(r)} \left(\frac{r}{r+\lambda t\theta_j}\right)^r \left(\frac{\lambda t\theta_j}{r+\lambda t\theta_j}\right)^{n_j} \quad (5)$$

where r is equal to $1/\alpha$. A suitable prior for α is given by a log-normal distribution similarly to σ_{c_i} to ensure α is strictly positive. Finally, in addition to estimating an independent σ_{c_i} per source, it is possible to alter the expectation $\lambda t\theta_j$ to estimate an independent expected number of events for each source, λ_{c_i}, where $\lambda = \Sigma\ \lambda_{c_i}$. The expected number of events at site j becomes

$$t\theta_j = t\pi\rho^2 \sum_{c_i=1}^{K} \lambda_{c_i} \cdot f_{\mathrm{BN}}\left(\mathbf{s}_j|\boldsymbol{\mu}_{c_i}, \boldsymbol{I}_2\sigma_{c_i}^2\right). \quad (6)$$

The likelihoods in Eqs. 4 and 5 can then be altered accordingly to accommodate independent λ_{c_i}.

The *silverblaze* package (Stevens and Verity 2021; see *Data Availability*) uses the likelihoods in Eqs. 4 and 5 combined with the priors in Eq. 1 to estimate the unknown parameters $\{\boldsymbol{\mu}_{c_i}\}$, $\{\sigma_{c_i}\}$ and $\{\lambda_{c_i}\}$ (for c_i in 1:K) in addition to α, under a negative binomial model, via MCMC methods using a combination of Metropolis-Hastings and Gibbs sampling. Details of the MCMC steps can be found in Appendix S2. A full list of model parameters can be found in Table 2.

Power analysis and model settings

We performed a Bayesian analogue of a traditional power analysis by simulating data from the Poisson model described in Eq. 1 and exploring the ability of the model to infer the true parameter values. For the validation of the Poisson model, we explored the parameter space similarly to Verity et al. (2014).

Source locations were generated uniformly at random from a longitudinal and latitudinal extent of −0.2 to 0.0 and 51.45 to 51.55 respectively. The spatial prior F was defined over a 100 × 100 grid whose extent matched the same values as the source locations plus a 25% margin at each limit (−0.25 to −0.05 and 51.425 to 51.575). This led to a spatial coverage of 345.53 km^2. The number of sources K ranged from one to five, the true value of σ_{c_i} was set to 1.5 km and the number of events N was Poisson distributed with rates 100, 1,000, and 10,000. For the power analysis, note that each source shared the same σ_{c_i} and λ_{c_i} (i.e., $\sigma_1 = \sigma_2 \ldots = \sigma_k$ and $\lambda_1 = \lambda_2 \ldots = \lambda_k$). This was chosen for simplicity given the study focused on the model's ability to estimate source locations in place of independent dispersal and expected number of events. Finally, the number of sentinel sites was set to 25, 100, or 400 and they were distributed over space either uniformly at random or as a grid.

Table 2. Parameters adopted in the methods sections.

Parameter	Definition
K	the true number of source locations
$\boldsymbol{\mu}_{c_i}$	the spatial location of source c_i in 1 to K
F	the prior on source locations
N	the number of events in a search area
λt	the rate of events in a search area in time t
(ζ, η)	the shape and rate of the gamma prior on λ
$\mathbf{x}_{c_i}$	the spatial location of event i, originating from source c_i
σ_{c_i} (km)	the bivariate normal's standard deviation centered on $\boldsymbol{\mu}_{c_i}$
(γ, δ)	the mean and variance of the lognormal prior on σ_{c_i}
m	the number of sentinel sites
ρ (km)	the sentinel site radius
$\mathbf{s}_j$	the spatial location of sentinel site j
n_j	the number of events encountered at sentinel site j
θ_j	the height of sentinel site j on the mixture of normal
$B_\rho(\mathbf{s}_j)$	the ball of radius ρ centered on sentinel site $\mathbf{s}_j$
f_{BN}	the density of the bivariate normal
α	the over-dispersion parameter governing count variance

To determine the correct number of source locations, the Poisson model ran seven times, in each case searching from one up to seven sources to allow for cases where the model overestimates K. The most suitable value of K was then chosen via the deviance information criterion (DIC; Spiegelhalter et al. 2014). The DIC is a metric for model comparison, used here to determine the best-fitting number of source locations. Parameter estimates for $\{\boldsymbol{\mu}_{c_i}\}$, $\{\sigma_{c_i}\}$, and $\{\lambda_{c_i}\}$ were then pulled for the value of K chosen by the DIC. These settings lead to 360 parameter combinations, each of which were repeated 100 times and results were averaged.

The log-normal prior on the dispersal σ_{c_i} was set as either tight (standard deviation of one) or wide (standard deviation of 100). The gamma prior on λ was set such that the mean was equal to the true rates (100, 1,000, or 10,000) and the standard deviation was either the true rate or 1/10th of the true rate.

The burn-in and sampling period for the MCMC chains were set to 5×10^4 iterations. Convergence of the MCMC chains during burn-in were determined by Geweke's metric for single MCMC chain convergence (Cowles and Carlin 1996). This was tested at multiples of 1×10^4 iterations during burn-in. To ensure healthy MCMC mixing, the new model utilized a Metropolis-Hastings coupling step (see Appendix S2; Atchadé et al. 2011).

The success of a geographic profiling model is measured via a source's hit score. This is defined as the area searched before finding a source divided by the total search area. Here our search strategy is defined by starting at the location with the highest value on the

geographic profile and working downwards. To summarize each simulation's hit scores we make use of the Gini coefficient, a metric developed in economics to describe the wealth distribution of a country across the population. In this context, we used it to describe the proportion of source locations discovered over area searched, where a coefficient of 1 corresponded to a perfect search strategy and 0.5 to a random search. Although we chose to represent the model's ability through the Gini coefficient, we could have equivalently represented this using the AUC, a metric commonly used in ecology. The Gini coefficient is calculated by scaling the AUC and was chosen for a clearer scale of model success ranging from 0 to 1 compared to the AUC that measures from 0.5 to 1 (Marcot 2012).

Mosquito surveillance data

Trap surveillance data from (Wilke et al. 2019) of the mosquito *Aedes aegypti* in Miami-Dade County, Florida were used to test the negative binomial model's ability to find breeding sites in ornamental bromeliads. Data consisted of 124 traps with encounters per trap ranging from 0 to 1033. A total of 94 traps contained *Aedes aegypti* and 30 did not. The average distance between an empty trap and its nearest positive trap was 55 m with a standard deviation of 77 m. There were 51 ornamental bromeliad patches that were checked for immature stages of mosquitoes where 30 contained *Aedes aegypti* larvae and 21 did not (Wilke et al. 2018). Here, we considered trap data recorded during 2017 to match the time period bromeliad patches were surveyed.

Model priors were set as follows. For source locations, the DPM model used a bivariate normal centered on the mean of the surveillance locations with standard deviation equal to the maximum distance between the data and mean (Verity et al. 2014). The final surface was then manipulated post-hoc to exclude the possibility of source locations in the sea using a shape file (Faulkner et al. 2018; shape file *available online*).[8] The negative binomial model used the same shape file for its prior on source locations where each cell's probability mass was uniform on land and zero in the sea. For the dispersal parameter σ_{c_i}, a diffuse prior was set for the DPM (mean of 2.5 and standard deviation of 10). The same hyperparameters were used for the negative binomial's prior on σ_{c_i} in addition to a tight prior (standard deviation of 1) to explore model behavior under different priors. These priors conform to previous studies placing *Ae. Aegypti* dispersal somewhere between 0 and 5 km (Service and Place 1997, Gorrochotegui-Escalante et al. 2000). For the negative binomial model, tight and diffuse log-normal priors were set for the expected number of events λ (means of 1×10^6 and standard deviations of 1×10^5 and 1×10^6). The prior on α was also log-normal with mean 1 and standard deviation 100.

To estimate the number of sources K, the negative binomial model was run 25 times, in each case searching for that specific number of sources, where the DIC was again utilized to pick the most suitable value of K to explain the data (Spiegelhalter et al. 2014). The DPM model used five sampling chains, each with a burn-in period of 5×10^2 iterations and a sampling period of 1×10^4 iterations. The negative binomial model ran for 5×10^4 burn-in and sampling iterations with convergence checked at each multiple of 1×10^4 iterations during burn-in (Cowles and Carlin 1996).

Software and data

The DPM, Poisson, and negative binomial models were developed in R and C++ and implemented in the *Rgeoprofile* (Verity and Le Comber 2021) and *Silverblaze* (Stevens and Verity 2021; see *Data Availability*) packages. In both cases, extensive documentation is available for installation and implementation. Furthermore, the R scripts used to run the analyses described in this manuscript in addition to the mosquito trap surveillance data and bromeliad breeding sites are available in Data S1.

Results

Gini coefficients

The results of the Bayesian power analysis can be seen in Table 3. There was a consistent decrease in power as we increased the number of sources but an increase in power given more sampling locations. Of the 360 parameter combinations, 278 (77%) reached a Gini coefficient of 0.9 or higher. Table 3 also shows that a uniform site configuration yielded a higher Gini coefficient more often than a random layout (134 of 180 cases). Additionally, tight priors on σ_{c_i} and λ in place of wide priors yielded higher Gini coefficients in 94 and 116 of 180 cases, respectively.

Parameter estimation

The new model was also tested on its ability to return the true number of source locations K, the true dispersal σ_{c_i} and finally, the expected number of events λ. The new model correctly fitted the true value of K in 57% of cases, it fitted within 1 of the true value in 76% of cases and within 2 in 88%.

The true σ_{c_i} value was set to 1.5 km. The model's average estimate for σ_{c_i} was 1.68 km (standard deviation of 0.94). The prior on σ_{c_i}, the prior on λ, the expected number of events λ, sampling strategy, number of sources, and number of sentinel sites all significantly affected the fitted value (ANOVA: σ_{c_i} prior $F_{1,\ 35{,}640} = 229.10$, $P < 2 \times 10^{-16}$; λ prior $F_{1,\ 35{,}640} = 7.20$,

[8] https://gis-mdc.opendata.arcgis.com/datasets/south-florida-region/

TABLE 3. Results of the Poisson model power analysis reported as mean Gini coefficient across replicates.

Site configuration, event rate (λ), and no. sites	Source 1				Source 2				Source 3				Source 4				Source 5			
	Tight λ prior		Wide λ prior		Tight λ prior		Wide λ prior		Tight λ prior		Wide λ prior		Tight λ prior		Wide λ prior		Tight λ prior		Wide λ prior	
	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior	Wide σ_{c_i} prior	Tight σ_{c_i} prior
Uniform																				
100																				
25	0.994	0.992	0.992	0.993	0.892	0.874	0.863	0.877	0.798	0.751	0.771	0.769	0.740	0.721	0.683	0.690	0.673	0.662	0.655	0.618
100	0.997	0.996	0.996	0.995	0.948	0.951	0.943	0.941	0.878	0.875	0.838	0.818	0.789	0.791	0.763	0.729	0.730	0.684	0.712	0.683
400	0.999	0.999	0.999	0.999	0.993	0.992	0.988	0.986	0.970	0.969	0.955	0.949	0.931	0.925	0.899	0.903	0.876	0.865	0.857	0.816
1,000																				
25	0.998	0.998	0.998	0.998	0.985	0.982	0.981	0.980	0.943	0.929	0.929	0.936	0.872	0.866	0.874	0.876	0.823	0.795	0.814	0.795
100	0.999	0.999	0.999	0.999	0.996	0.997	0.996	0.996	0.989	0.988	0.982	0.984	0.973	0.971	0.968	0.954	0.946	0.935	0.934	0.918
400	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.997	0.997	0.996	0.996	0.993	0.993	0.992	0.992	0.986	0.988	0.980	0.983
10,000																				
25	0.999	0.999	0.999	0.999	0.998	0.998	0.997	0.996	0.995	0.993	0.994	0.989	0.985	0.985	0.984	0.977	0.969	0.963	0.954	0.955
100	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.997	0.998	0.997	0.996	0.995	0.996	0.993	0.994
400	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998
Random																				
100																				
25	0.994	0.994	0.991	0.992	0.904	0.895	0.899	0.868	0.822	0.807	0.773	0.783	0.715	0.750	0.702	0.700	0.689	0.677	0.660	0.664
100	0.996	0.994	0.995	0.994	0.950	0.942	0.935	0.928	0.853	0.855	0.816	0.829	0.775	0.768	0.747	0.755	0.721	0.722	0.693	0.692
400	0.997	0.996	0.998	0.997	0.984	0.990	0.984	0.984	0.959	0.946	0.943	0.939	0.904	0.907	0.894	0.872	0.847	0.845	0.845	0.810
1,000																				
25	0.998	0.996	0.997	0.998	0.969	0.964	0.969	0.965	0.908	0.910	0.901	0.918	0.853	0.845	0.833	0.821	0.814	0.809	0.744	0.766
100	0.999	0.999	0.999	0.999	0.994	0.995	0.995	0.993	0.986	0.987	0.980	0.981	0.968	0.958	0.949	0.957	0.941	0.931	0.906	0.902
400	0.999	0.999	0.999	0.999	0.998	0.996	0.997	0.998	0.994	0.995	0.995	0.993	0.989	0.990	0.991	0.989	0.981	0.977	0.978	0.978
10,000																				
25	0.999	0.998	0.999	0.999	0.997	0.995	0.988	0.989	0.984	0.982	0.982	0.976	0.966	0.965	0.961	0.950	0.938	0.937	0.926	0.912
100	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.999	0.998	0.998	0.997	0.996	0.995	0.995	0.992	0.991	0.991	0.990
400	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.999	0.999	0.998	0.998	0.998	0.999	0.998	0.996	0.998	0.997	0.996

$P = 0.01$; expected events $F_{2,\ 35,640} = 2841.77$, $P < 2 \times 10^{-16}$; sampling strategy $F_{1,\ 35,640} = 224.98$, $P < 2 \times 10^{-16}$; sources $F_{4,\ 35,640} = 394.54$, $P < 2 \times 10^{-16}$; sentinel sites $F_{2,\ 35,640} = 735.24$, $P < 2 \times 10^{-16}$). Of all the interactions, the true expected number of events remained the strongest variable that affected the fitted value of σ_{c_i}.

True λ values were set to 100, 1,000, and 10,000. The model's average estimates for λ were 118, 1,094, and 10,501 (with standard deviations of 34, 274, and 1,856, respectively). Of the same list of variables, all significantly affected the fitted value for the expected number of events, with the exception of the number of sentinel sites (ANOVA: σ_{c_i} prior $F_{1,\ 35,640} = 332.40$, $P < 2 \times 10^{-16}$; λ prior $F_{1,\ 35,640} = 1105.00$, $P < 2 \times 10^{-16}$; expected events $F_{2,\ 35,640} = 4.092 \times 10^5$, $P < 2 \times 10^{-16}$; sampling strategy $F_{1,\ 35,640} = 98.27$, $P < 2 \times 10^{-16}$; sources $F_{4,\ 35,640} = 18.72$, $P = 2 \times 10^{-15}$; sentinel sites $F_{2,\ 35,640} = 268.70$, $P < 2 \times 10^{-16}$). In the case of interactions, the strongest variable that affected the fitted expected number of events was the true expected number of events.

Mosquito surveillance data

The mosquito surveillance data and bromeliad patches can be seen alongside the geographic profiles created by the negative binomial and DPM models in Fig. 1a and b. The DPM model determined 91 clusters best described the data. Within the negative binomial model, the DIC determined varying cluster numbers (between 2 and 18) dependent on the choice of parameter priors (Fig. 2). Hit score percentages for the DPM model ranged from 0.13% to 41.64% with an average of 11.12%. The negative binomial model's hit scores percentages ranged from 1.95% to 69.00% with an average of 21.27%. Under informative priors the negative binomial model returned a dispersal σ_{c_i} value between 1.41 and 7.03 km (95% credible interval) whereas under less informative priors estimates reached up to 22 km. Comparatively, the DPM model estimated σ_{c_i} between 9 and 10 m. The total expected population density of *Aedes aegypti* was estimated between 3.64 to 28.28 million for 2017. The over-dispersion parameter α was consistently estimated between 2.40 and 4.35.

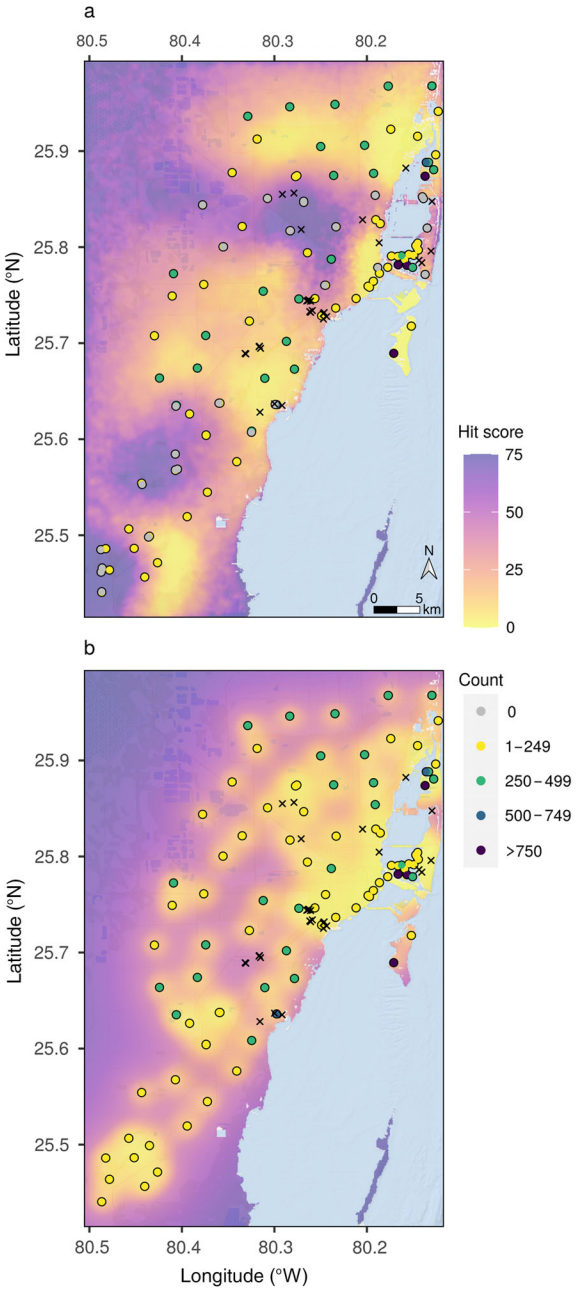

FIG. 1. The geographic profiles in Miami-Dade County, Florida, USA created by (a) the negative binomial model via the 2017 mosquito count data under informative priors ($K = 14$) and (b) the DPM model via repeat point-pattern data ($K = 91$). Locations of bromeliad breeding sites are marked with a cross (Wilke et al. 2018, 2019). Given the proximity between positive and empty traps, some positive traps are only visible in Fig. 1b.

DISCUSSION

In this paper we have constructed and validated a new geographic profiling model that can distinguish between an absence of evidence and evidence of absence. This was done by taking as input count data into the model's likelihood of which can consist of locations associated with no encounters.

Accounting for different information can lead to different search strategies. Sentinel sites with no encounters drew us away from common search practices such as looking near the spatial mean of observed data, a method that is only effective when searching for a single source location (Stevenson et al. 2012, Verity et al. 2014). In addition, this new information drew search priority away from those areas containing no encounters to those with no information at all, compared to the DPM model, where these areas were treated equally. An assumption when using the DPM model is that perfect

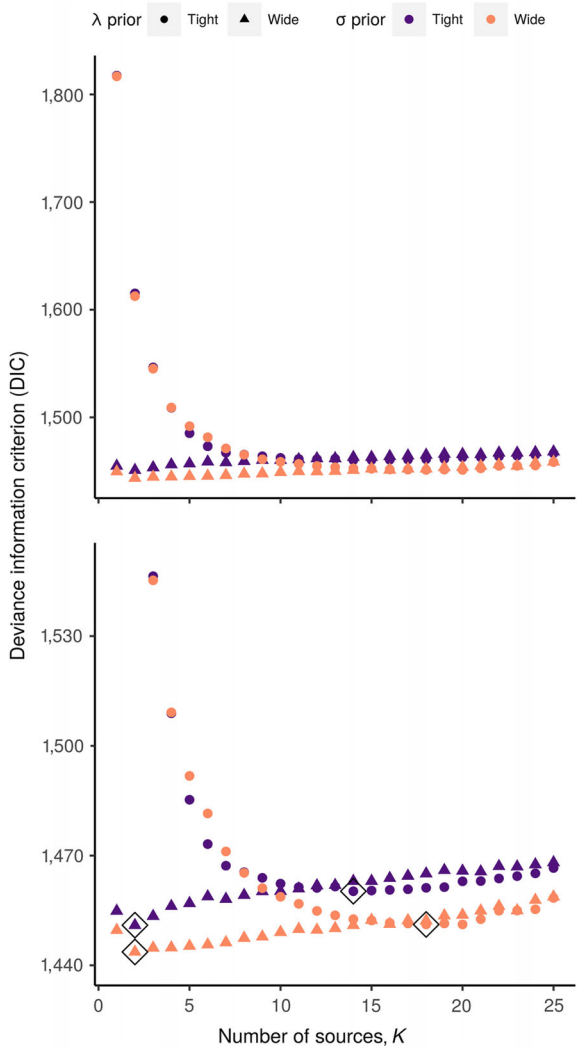

FIG. 2. The deviance information criterion (DIC) for each negative binomial model searching for K source locations under different priors. The most suitable number of source locations based on prior combination are marked with a diamond. Full DIC values are displayed in the top panel with a zoomed version on the bottom.

observations are made, meaning all events that occur will be seen. This assumption is valid in studies where the exact locations of events are recorded (Faulkner et al. 2015, Smith et al. 2015, Struebig et al. 2018) but is less suitable in those that adopt a sampling strategy using sentinel sites (Faulkner et al. 2016).

We have shown via a power analysis and real-world case study that the new model can estimate a variety of parameters common to geographic profiling in addition to new ones. It accurately estimated source locations, dispersal σ_{c_i} and the number of source locations, K, in addition to the newly fitted expected number of events, λ, and over-dispersion parameter, α. Finally, it allows for cases where σ_{c_i}, and λ vary from source to source.

The new model was able to identify source locations efficiently, as was reflected in consistently high Gini coefficients across parameter combinations. Average Gini coefficients never fell below 0.5, the value associated with a random search strategy.

Estimating the number of sources in the new model was less straightforward than in the DPM model. A major strength of the latter is that it did not require us to specify the number of source locations in advance. In the new model, we ran the algorithm many times and used the deviance information criterion (DIC) to find the most appropriate number of source locations. This process produced accurate results for simulated data but was shown to produce different results dependent on prior choice in the real-world case study. Here, combinations of diffuse and informative priors indicated suitable K values at 2, 14, or 18 (Fig. 2). Although a K value of 2 corresponded to the best DIC value, estimates of σ_{c_i} in this case were up to 22 km. For a K value of 14, estimates were much more sensible. The DPM model estimated σ_{c_i} between 9 and 10 m. In both cases, each model's estimate for σ_{c_i} contradicted our prior beliefs built from our biological understanding of *Ae. Aegypti* dispersal. We therefore suggest careful consideration be taken when building priors and advice from field experts and collaborators is sought.

It would be naïve to assume the number of sources fitted by either models or the known number of breeding sites reflects the true number of sources, of which could consist of any body of stagnant water (Ramasamy et al. 2011). Given the ground truth about the true number of sources is unknown, there is no way of evaluating the hit scores of these hypothetical locations. We therefore suggest that the number of source locations fitted by either model play the role of a lower bound on the true value of K. We also suggest future work could focus on migrating the new model to a nonparametric framework, similarly to the DPM model, in place of estimating K by running the model multiple times.

In this new model, we derived an expectation for the number of events at each sentinel site. This was dependent on the site radius ρ, expected number of events in our search area λ, time the sentinel is left open (susceptible to events), and the site's spatial location with respect to sources.

The sentinel site radius ρ was kept constant throughout our analyses to ensure the approximation in Eq. 2 was not erroneous (see Appendix S1). In our model, an event was encountered by a single sentinel site only. This was based upon whether an event fell within a site's radius. Should an event be encountered by two sites, then we must have observed it at two distinct points in time. The effect of the site radius is like that of time; the larger the radius the more events expected at each site. Here, the effect of time was not explored, rather its units were set to the time interval in which data were collected. As suggested in many studies (Rossmo 2000, Raine et al. 2009, Santosuosso and Papini 2018), a more accurate

geographic profiling model is one that considers temporal variability in the data to draw its inferences.

A sentinel site that encounters at least one event is indicative of the presence of, for example, an invasive species. The opposite, however, is not necessarily true for a site that encounters nothing. If a sentinel site yields no encounters, then either an event is not present in that area or, it is, but the sentinel site failed to observe it. In this study if an event fell within a sentinel site's radius then it was immediately encountered by that site. Detection probabilities are not always one and future studies may investigate relaxing this condition. Furthermore, we could adapt the observation model so that encounters are not governed by a site radius, such as in (Chandler and Royle 2013).

Collecting count data is common in ecology, for example in spatially explicit capture–recapture models and site occupancy models (MacKenzie et al. 2002, Kéry et al. 2011, Royle et al. 2011, Chandler and Royle 2013). The primary purpose of these models is to estimate abundance, rather than, as here, the location of sources. Spatially explicit capture–recapture models do treat these source locations (known as "activity centers") as a latent variable but make differing assumptions about their numbers. Instead of assuming each encountered event is associated with a unique source, geographic profiling aims to partition the count data into clusters and finds the source location associated with each cluster. The aim of this study was to build a model that estimated source locations using count data, so the architecture of the new model was built from the point of view of historical geographic profiling models that consistently focus on estimating this parameter.

In addition to count data, it is entirely possible for the new model to utilize pseudo-absences in its inference process (Barbet-Massin et al. 2012). By replacing unsampled locations with pseudo-absences we would expect the model to focus search priority entirely on locations with positive data. However, this could be accomplished by a suitably informed Bayesian prior on source locations, such as the Miami-Dade coastline shape file that was used to ignore locations in the sea. Comparing the utility between a Bayesian prior and a set of pseudo-absences both derived from habitat suitability was not tested here but could be explored in future work. Geographic heterogeneities have been considered in previous geographic profiling models such as (Mohler and Short 2012). This is however, the first time we see such information accounted for in a geographic profiling model that can also estimate multiple numbers of sources.

Conclusions

Our analyses and results have shown that a geographic profiling model that utilizes count data can alter search strategies when intervening in cases of species invasion, outbreak of infection or crime by making the distinction between evidence of absences in data and an absence of evidence. In doing so, search strategies produced move priority away from those locations containing absences to those containing no information at all; a substantial change over existing models that treat these areas with equal search priority. Additionally, the new model introduces the ability to estimate spatial dispersal and expected population size unique to each source location as well as the flexibility to a user to implement any spatial prior desired. Different models should be used in differing circumstances dependent on the type of data to hand. The DPM model should be used when data are in point-pattern form (each location is associated with a single instance of crime, an invasive species, or disease) and the new model should be chosen when we have a list of sentinel site locations and associated counts (bioacoustics monitors, cameras, or pitfall traps).

Acknowledgments

This research utilized Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT, http://doi.org/10.5281/zenodo.438045. This work was supported by the Natural Environmental Research Council (grant number NE/L002485/1). This research was supported by the CDC (https://www.cdc.gov/) grant 1U01CK000510-05: Southeastern Regional Center of Excellence in Vector-Borne Diseases: The Gateway Program. CDC had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. We would like to thank two reviewers for contributing invaluable comments to an earlier version of this manuscript. We also thank Joe Nichols, whose question about football grounds provided the original impetus for this study. Author contributions: M. C. A. Stevens, R. Verity, R. A. Nichols, and S. Le Comber were responsible for conceptualizing and building the methodology as well as writing and reviewing the manuscript; A. B. B. Wilke, J. C. Beier, C. Vasquez, and W. D. Petrie were responsible for data collection and curation; M. C. A. Stevens ran the formal analysis; M. C. A. Stevens and R. Verity built the software for the manuscript. All authors reviewed and approved the final version of the manuscript.

Literature Cited

Atchadé, Y. F., G. O. Roberts, and J. S. Rosenthal. 2011. Towards optimal scaling of metropolis-coupled Markov chain Monte Carlo. Statistics and Computing 21:555–568.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: How, where and how many? Methods in Ecology and Evolution 3:327–338.

Borchers, D. 2012. A non-technical overview of spatially explicit capture-recapture models. Journal of Ornithology 152:435–444.

Borchers, D. L., and M. G. Efford. 2008. Spatially explicit maximum likelihood methods for capture-recapture studies. Biometrics 64:377–385.

Butkovic, A., S. Mrdovic, S. Uludag, and A. Tanovic. 2018. Geographic profiling for serial cybercrime investigation. Digital Investigation 28:176–182.

Cerri, J., E. Mori, R. Zozzoli, A. Gigliotti, A. Chirco, and S. Bertolino. 2020. Managing invasive Siberian chipmunks Eutamias sibiricus in Italy: a matter of attitudes and risk of dispersal. Biological Invasions 22:603–616. http://dx.doi.org/10.1007/s10530-019-02115-5

Chandler, R. B., and A. J. Royle. 2013. Spatially explicit models for inference about density in unmarked or partially marked populations. Annals of Applied Statistics 7:936–954.

Cowles, M. K., and B. P. Carlin. 1996. Markov Chain Monte Carlo Convergence diagnostics: a comparative review. Journal of the American Statistical Association 91:883–904.

Doney, R. 1990. The aftermath of the Yorkshire Ripper: the response of the United Kingdom Police Service. Pages 95–112 *in* S. Egger, editor. Serial murder: an elusive phenomenon. Praeger, New York, New York, USA.

Efford, M. 2004. Density estimation in live-trapping studies. Oikos 106:598–610.

Faulkner, S. C., M. C. A. Stevens, S. S. Romañach, P. A. Lindsey, and S. C. Le Comber. 2018. A spatial approach to combatting wildlife crime. Conservation Biology 32:685–693.

Faulkner, S. C., M. D. Stevenson, R. Verity, A. H. Mustari, S. Semple, D. G. Tosh, and S. C. Le Comber. 2015. Using geographic profiling to locate elusive nocturnal animals: A case study with spectral tarsiers. Journal of Zoology 295:261–268.

Faulkner, S. C., R. Verity, D. Roberts, S. S. Roy, P. A. Robertson, M. D. Stevenson, and S. C. Le Comber. 2016. Using geographic profiling to compare the value of sightings vs trap data in a biological invasion. Diversity and Distributions 23:104–112.

Gorrochotegui-Escalante, N., M. De Lourdes Munoz, I. Fernandez-Salas, B. J. Beaty, and W. C. IV Black. 2000. Genetic isolation by distance among *Aedes aegypti* populations along the northeastern coast of Mexico. American Journal of Tropical Medicine and Hygiene 62:200–209.

Hayes, E. B. 2009. Zika virus outside Africa. Emerging Infectious Diseases 15:1347–1350.

Heald, O. J. N., C. Fraticelli, S. E. Cox, M. C. A. Stevens, S. C. Faulkner, T. M. Blackburn, and S. C. Le Comber. 2019. Understanding the origins of the ring-necked parakeet in the UK. Journal of Zoology 312:1–11.

Hennessey, M., M. Fischer, and J. Staples. 2016. Zika virus spreads to new areas—region of the Americas, May 2015–January 2016. American Journal of Transplantation 16:1031–1034.

Kéry, M., B. Gardner, T. Stoeckle, D. Weber, and J. A. Royle. 2011. Use of spatial capture-recapture modeling and DNA data to estimate densities of elusive animals. Conservation Biology 25:356–364.

Le Comber, S. C., B. Nicholls, D. K. Rossmo, and P. A. Racey. 2006. Geographic profiling and animal foraging. Journal of Theoretical Biology 240:233–240.

Le Comber, S. C., D. K. Rossmo, A. N. Hassan, D. O. Fuller, and J. C. Beier. 2011. Geographic profiling as a novel spatial tool for targeting infectious disease control. International Journal of Health Geographics 10:1–8.

Lindén, A., and S. Mäntyniemi. 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92:1414–1421.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, A. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

Marcot, B. G. 2012. Metrics for evaluating performance and uncertainty of Bayesian network models. Ecological Modelling 230:50–62.

Martin, R. A., D. K. Rossmo, and N. Hammerschlag. 2009. Hunting patterns and geographic profiling of white shark predation. Journal of Zoology 279:111–118.

Mohler, G. O., and M. B. Short. 2012. Geographic profiling from kinetic models of criminal behavior. SIAM Journal on Applied Mathematics 72:163–180.

O'Leary, M. 2009. The mathematics of geographic profiling. Journal of Investigative Psychology and Offender Profiling 6:253–265.

O'Leary, M. 2010. Implementing a Bayesian approach to criminal geographic profiling. Pages 1–8 *in* 1st International Conference and Exhibition on Computing for Geospatial Research & Application. Association for Computing Machinery, Washington, D.C., USA.

Papini, A., S. Mosti, and U. Santosuosso. 2013. Tracking the origin of the invading Caulerpa (Caulerpales, Chlorophyta) with Geographic Profiling, a criminological technique for a killer alga. Biological Invasions 15:1613–1621.

Raine, N. E., D. K. Rossmo, and S. C. Le Comber. 2009. Geographic profiling applied to testing models of bumble-bee foraging. Journal of the Royal Society 6:307–319.

Ramasamy, R., S. N. Surendran, P. J. Jude, S. Dharshini, and M. Vinobaba. 2011. Larval development of *Aedes aegypti* and *Aedes albopictus* in peri-urban brackish water and its implications for transmission of arboviral diseases. PLoS Neglected Tropical Diseases 5:e1369.

Rossmo, D. K. 1993. A methodological model. American Journal of Criminal Justice 17:1–21.

Rossmo, D. K. 2000. Geographic profiling. CRC Press, Boca Raton, Florida, USA.

Rossmo, D. K. 2012. Recent developments in geographic profiling. Policing 6:144–150.

Rossmo, D. K., H. Lutermann, M. D. Stevenson, and S. C. Le Comber. 2014. Geographic profiling in Nazi Berlin: fact and fiction. Geospatial Intelligence Review 12:44–57.

Rossmo, D. K., and R. Routledge. 1990. Estimating the size of criminal populations. Journal of Quantitative Criminology 6:293–314.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A., A. K. Fuller, and C. Sutherland. 2018. Unifying population and landscape ecology with spatial capture–recapture. Ecography 41:444–456.

Royle, J. A., A. J. Magoun, B. Gardner, P. Valkenburg, and R. E. Lowell. 2011. Density estimation in a wolverine population using spatial capture-recapture models. Journal of Wildlife Management 75:604–611.

Santosuosso, U., and A. Papini. 2018. Geo-Profiling: beyond the current limits. A preliminary study of mathematical methods to improve the monitoring of invasive species. Russian Journal of Ecology 49:346–354.

Service, M. W., and P. Place. 1997. Mosquito (Diptera: Culicidae) dispersal—the long and short of it. Journal of Medical Entomology 34:579–588.

Smith, C. M., S. H. Downs, A. Mitchell, A. C. Hayward, H. Fry, and S. C. L. Comber. 2015. Spatial targeting for bovine tuberculosis control: Can the locations of infected cattle be used to find infected badgers? PLoS ONE 10:1–14.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2014. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B 76:485–493.

Stevens, M., and R. Verity. 2021. Michael-Stevens-27/silverblaze: First Release (Version v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4569035

Stevenson, M. D., D. K. Rossmo, R. J. Knell, and S. C. Le Comber. 2012. Geographic profiling as a novel spatial tool for targeting the control of invasive species. Ecography 35:704–715.

Struebig, M. J., et al. 2018. Addressing human-tiger conflict using socio-ecological information on tolerance and risk. Nature Communications 9:3455.

Ver Hoef, J. M., and P. L. Boveng. 2007. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88:2766–2772.

Verity, R., and S. Le Comber. 2021. bobverity/Rgeoprofile: molecular ecology release (Version mol_ecol). Zenodo. https://doi.org/10.5281/zenodo.4608713

Verity, R., M. D. Stevenson, D. K. Rossmo, R. A. Nichols, and S. C. Le Comber. 2014. Spatial targeting of infectious disease control: identifying multiple, unknown sources. Methods in Ecology and Evolution 5:647–655.

Wilke, A. B. B., C. Vasquez, P. J. Mauriello, and J. C. Beier. 2018. Ornamental bromeliads of Miami-Dade County, Florida are important breeding sites for *Aedes aegypti* (Diptera: Culicidae). Parasites & Vectors 11:1–7.

Wilke, A. B. B., C. Vasquez, J. Medina, A. Carvajal, W. Petrie, and J. C. Beier. 2019. Community composition and year-round abundance of vector species of mosquitoes make Miami-Dade County, Florida a receptive gateway for arbovirus entry to the United States. Scientific Reports 9:8733.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2329/full

Data Availability

The DPM, Poisson, and negative binomial models were developed in R and C++ and implemented in the *Rgeoprofile* (Verity and Le Comber 2021; https://doi.org/10.5281/zenodo.4608713) and *Silverblaze* (Stevens and Verity 2021; http://doi.org/10.5281/zenodo.4569035) packages.