Received ; | Revised ; | Accepted

DOI: xxx/xxxx

ARTICLE TYPE

Modelling disease dynamics from spatially explicit capture-recapture data

Fabian R. Ketwaroo[1] | Eleni Matechou[1] | Matthew Silk[2] | Richard Delahay[3]

[1]School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK
[2]CEFE, University of Montpellier, CNRS, EPHE, IRD, Montpellier, France
Centre for Ecology and Conservation, University of Exeter, Penryn, UK
[3]Wildlife Disease Ecology Team, Central Science Laboratory , Sand Hutton, York YO41 1LZ, UK

Correspondence
Fabian Ketwaroo. Email: fk231@kent.ac.uk

Summary

One of the main aims of wildlife disease ecology is to identify how disease dynamics vary in space and time and as a function of population density. However, monitoring spatio-temporal and density-dependent disease dynamics in the wild is challenging because the observation process is error-prone, which means that individuals, their disease status and their spatial locations are unobservable, or only imperfectly observed. In this paper, we develop a novel spatially explicit capture-recapture (SCR) model motivated by an SCR data set on European badgers (*Meles Meles*), naturally infected with bovine tuberculosis (*Mycobacterium bovis*). Our model accounts for the observation process of individuals as a function of their latent activity centres, and for their imperfectly observed disease status and its effect on demographic rates and behaviour. This framework has the advantage of simultaneously modelling population demographics and disease dynamics within a spatial context. It can therefore generate estimates of critical parameters such as population size; local and global density by disease status and hence spatially-explicit disease prevalence; disease transmission probabilities as functions of local or global population density; and demographic rates as functions of disease status. Our findings suggest that infected badgers have lower survival probability but larger home range areas than uninfected badgers, and we found no clear evidence of density-dependence in disease transmission. We also present an extensive simulation study, considering different scenarios of disease transmission within the population, and our findings highlight the importance of accounting for spatial variation in disease transmission and individual disease status when these affect demographic rates. Collectively these results show our new model enables a better understanding of how wildlife disease dynamics are linked to population demographics within a spatio-temporal context.

KEYWORDS:
Density-dependent transmission, European badger, endemic disease, population density

1 | INTRODUCTION

Understanding relationships between the dynamics of populations and their infectious pathogens is a key aim of wildlife disease ecology. Infectious diseases can directly influence population density through impacts on demographic vital rates (Manlove, Cassirer, Cross, Plowright, & Hudson 2016; McCallum et al. 2007; Vredenburg, Knapp, Tunstall, & Briggs 2010), but in turn disease dynamics can vary in space and time and with population density. Quantifying how pathogen transmission varies with population density is important because of its implication for the conservation and management of wildlife populations (McCallum 2016; Silk et al. 2019). While the potential importance of host density for epidemiological dynamics has long been known (Krkošek 2010; O'Neill, White, Gortázar, & Ruiz-Fons 2023), it is now recognized that aspects of spatial and social behaviour (e.g. Silk & Fefferman 2021) can mean that pathogen transmission may covary with host population density across a continuum from density-dependent (transmission increases with population density) to frequency-dependent (transmission is independent of population density) (Hopkins, Fleming-Davies, Belden, & Wojdak 2020).

There are two main analytical challenges associated with quantifying the relationship between host density and disease transmission in wildlife populations. First is the difficulty of estimating population size, and hence population density, in wild populations, where not every individual is captured or observed over time. The second challenge is associated with the detection of infection in wild hosts owing to limitations in the sensitivity and specificity of diagnostic tests (Choquet, Carrié, Chambert, & Boulinier 2013; Drewe, Tomlinson, Walker, & Delahay 2010; Enøe, Georgiadis, & Johnson 2000).

Capture-recapture (CR) models have been one of the main tools developed to estimate the population size of wild animals. Traditional CR models essentially represent "fish bowl" sampling, that is, a system that is unconnected to the spatial structure of the population. These models do not account for the spatial nature of sampling nor the spatial distribution of individuals (J. A. Royle, Fuller, & Sutherland 2018). Consequently, they do not allow for the study of many vital spatial population processes, such as density, movement and dispersal of individuals. This weakness of CR models has been overcome by the development of spatially-explicit capture-recapture (SCR) models (Borchers & Efford 2008; Efford 2004). SCR models consider the collection of individuals in a population as a latent point process of where an individual moves (home range), which is centred on their individual activity centre (AC), distributed within some region of interest.

In SCR data, individual ACs are unknown and thus are considered latent variables in corresponding models. SCR models can be fitted in a classical framework, where the ACs are marginalised from the likelihood by integration (Borchers & Efford 2008), or in a Bayesian framework, where the ACs are explicitly estimated along with other unknown parameters and random variables using Markov chain Monte Carlo (MCMC) methods (J. A. Royle & Young 2008). Once inferred, the ACs can be used to estimate spatial population processes such as density, which is the number of ACs per unit area of the region of interest. Additionally, conditional on the latent ACs, the probability of observing or encountering an individual is modelled as a function

of the distance between the individual's AC and the location of each trap. Consequently, SCR models take into consideration the spatial nature of sampling as well as spatial distribution of individual ACs to allow for the study of spatial population processes, which is arguably equally important to the study of demographic population rates, with the formal link between state model and observational model allowing for better inference on the former and more robust accommodation of the latter (Sutherland, Royle, & Linden 2019).

However, existing SCR models do not currently accommodate additional data on the disease status of captured individuals. These data typically consist of diagnostic test results that are also prone to error and hence are only an imperfect observation of an individual's disease status. Additionally, as is always the case in CR data, these individual-level diagnostics are only available for the particular time point when an animal is captured, and are missing for all other times and for individuals that have never been caught.

In the present study, we develop a novel SCR model that accounts for the observation process of individuals, as well as their imperfectly observed disease status. Our new modelling framework allows the simultaneous modelling of population demographics and disease dynamics within a spatiotemporal context. This makes it possible to simultaneously test hypotheses related to spatial and density-related variation in disease transmission alongside examining variation in survival and individual capture probabilities as a function of individual (latent) disease state.

We perform an extensive simulation study to assess model performance for a number of scenarios. Our results demonstrate the quality of inference in our proposed model but also reveal the requirements in terms of data size and effect size to have sufficient power to identify density-dependence in disease transmission. We highlight that when demographic rates are dependent upon individual disease status, existing SCR models, which do not account for that dependence, yield substantially biased estimates of population density.

We then fit our model to a motivating case study of European badgers (*Meles meles*), naturally infected with bovine tuberculosis (*Mycobacterium bovis*) at Woodchester Park in Gloucestershire, UK (R. J. Delahay et al. 2013; McDonald, Robertson, & Silk 2018), revealing important aspects of the epidemiological dynamics. Our model reveals that infected individuals have a lower survival probability and a larger home range. We also infer that despite a more recent reduction in population size, the prevalence of infection remained constant suggesting no clear evidence of density-dependence in disease transmission in the population during the recent study period.

The paper is structured as follows: in Section 2 we describe the case study that motivated the work in this paper, in Section 3 we introduce the new model and discuss our inference approach, while Sections 4 and 5 present simulation and case study results, respectively. Section 6 discusses the results from our simulations and case study in the context of wildlife disease ecology, and suggests directions for future work.

2 | DATA COLLECTION AND PROCESSING

The Woodchester Park study area is located on the Cotswold limestone escarpment in Gloucestershire, South-west England. Over an area of approximately 7 km^2 the resident badger population has been monitored in a consistent manner since 1981 (McDonald et al. 2018). The majority of the study area comprises mixed woodland, grassland, and arable farmland (R. Delahay, Carter, Forrester, Mitchell, & Cheeseman 2006). Badger population density is relatively high (Rogers, Cheeseman, Mallinson, & Clifton-Hadley 1997) and their social groups occupy more or less contiguous territories throughout the study area (R. Delahay et al. 2006), each of which is typically associated with one main sett (underground burrow system).

The badger population is monitored by CR sampling, which enables the collection of demographic and epidemiological data. The study area has been divided into three zones of approximately equal size, each of which is trapped four times per year from May to January inclusive, with a suspension from February to April to avoid catching dependent cubs and their mothers (Woodroffe et al. 2006). To determine which setts are active and how many traps to deploy, a sett activity survey is conducted in each zone before each trapping event. At each active sett, more traps are deployed than are expected to be required (i.e. saturation trapping).

Box traps constructed of steel mesh are dug into the substrate close to each active sett and baited with peanuts for four to eight days to habituate badgers to their presence (Cheeseman, CL, et al. 1982). On the last day of baiting, the traps are set for two consecutive nights and are checked on the following mornings. Once captured, newly caught badgers are aged, sexed and permanently marked with a unique ID tattoo on the abdomen (Cheeseman et al. 1982), with their weight, body condition, body length, reproductive status and tooth wear recorded at this and every subsequent capture event.

Once captured, three tests are currently used to test for *M.bovis*: interferon-gamma immunoassay (Ifn, Dalley et al. 2008), used since 2006 to detect a cell-mediated immune response, Dual Path Platform test (DPP®, Chembio.inc), used since 2014 to test for antibodies (Ashford et al. 2020) and elective microbiological culture of clinical samples (Cul, Gallagher & Horwill 1977) used since 1976. Ifn and DPP® use blood samples whilst Cul is carried out on samples of sputum, faeces, urine, and swabs of abscesses and wounds. Each test is imperfect, resulting in false positive and false negative errors, making it difficult to infer an individual's disease state from the tests alone (Drewe et al. 2010).

Following examination and collection of diagnostic samples, badgers caught during the first night of trapping are held overnight and released the following morning to prevent them from being re-captured on the second trapping night. Badgers caught during the second night are released the following day. Badgers are released where captured, following a period of recovery and subject to welfare assessment.

3 | MODEL

SCR models assume that a population of $i = 1, \dots, N_t$ individuals are monitored at $t = 1, \dots, T$ sampling occasions and $j = 1, \dots, J$ sampling locations, and each individual has an associated spatial location within a spatial domain (S), representing its AC $s_{i,t}$, a two-dimensional spatial coordinate. The collection of ACs can be thought of as a statistical spatial point pattern that describes how individuals are distributed within S. This statistical point process is often referred to as the state model. Here, we define our model in a Bayesian framework using data augmentation (DA, J. A. Royle & Dorazio 2012) and let $i = 1, \dots, M$ be "pseudo-individuals" that potentially could belong to N_t. In what follows, $i = 1, \dots, M$ indexes individuals, $t = 1, \dots, T$ sampling occasions and $j = 1 \dots, J$ sampling locations.

Our model has two key latent states: presence, $z_{i,t}$, and disease status, $d_{i,t}$, defined as

$$z_{i,t} = \begin{cases} 1 & \text{alive on sampling occasion t} \\ 0 & \text{unrecruited/dead on sampling occasion t.} \end{cases}$$

$$d_{i,t}|z_{i,t} = 1 = \begin{cases} 1 & \text{infected and alive on sampling occasion t} \\ 0 & \text{uninfected and alive on sampling occasion t.} \end{cases}$$

We assume that individual ACs do not change over time by modelling

$$s_{i,t} = s_i \sim \text{Uniform}(S)\ \forall i, t$$

Naturally, observation $y_{i,j,t}$ (equal to 1 if individual i was caught on sampling occasion t and sampling location j and 0 otherwise) depends on the corresponding presence state and AC

$$y_{i,j,t}|z_{i,t} \sim \text{Bernoulli}(p(x_j, s_i)z_{i,t})$$

where $p(x_j, s_i)$, which corresponds to the probability that individual i is caught on sampling location j, conditional on being present, is modelled by the half-normal function (Efford 2004)

$$p(x_j, s_i) = p_{0_{d_{i,t}}} \exp\Big(-\frac{1}{2\sigma^2_{d_{i,t}}} \parallel x_j - s_i \parallel^2 \Big)$$

where x_j is the two-dimensional spatial coordinate location of the j^{th} sampling location, $p_{0_{d_{i,t}}}$ is the baseline capture probability and $\sigma_{d_{i,t}}$ represents the rate at which capture probability declines as Euclidean distance from the AC increases. We model both

of these parameters dependent on the individual disease status at occasion t, using a logistic regression for p_0 and a log-linear model for σ, both of which have the binary disease status as the only covariate, allowing disease status to potentially affect behaviour in terms of space use.

Finally, we model the result of test Q, $\omega_{i,t}^{Q}$, for individual i on occasion t, conditional on their disease status, as a Bernoulli($\omega_{i,t}^{Q}$) random variable, with

$$\omega_{i,t}^{Q}|d_{i,t} = \begin{cases} 1 - q_{00}^{Q}, & d_{i,t} = 0 \\ q_{11}^{Q}, & d_{i,t} = 1 \end{cases}$$

where we refer to the probability of a true positive result by test Q as q_{11}^{Q} (sensitivity of test Q) and to the corresponding probability of a true negative result as q_{00}^{Q} (specificity of test Q), and $Q \in \{\text{DPP}^{®}, \text{Ifn}, \text{Cul}\}$. Following Buzdugan, Vergne, Grosbois, Delahay, and Drewe (2017), we assume independence between tests and hence define the joint distribution of the three test results as the product of the marginal Bernoulli distributions. The sensitivity and specificity of each test are inferred parameters, thus, enabling the diagnostic accuracy of each test to be evaluated. This formulation accounts for imperfect tests and enables a higher diagnostic accuracy than single test use (Drewe et al. 2010).

We model the transition between latent states accordingly, so that at $t = 1$

$$z_{i,1} \sim \text{Bernoulli}(\gamma_1)$$

$$d_{i,1}|z_{i,1} \sim \text{Bernoulli}(z_{i,1}\delta_I)$$

where γ_1 is the recruitment probability that a "pseudo-individual" is in the population at the start of the study and δ_I is the probability of being infected at the start of the study. For $t \geq 2$,

$$z_{i,t} \sim \text{Bernoulli}(\phi_{d_{i,t-1}} z_{i,t-1} + \gamma_t \alpha_{i,t})$$

$$d_{i,t} \sim \text{Bernoulli}(z_{i,t}\left[d_{i,t-1} + \left\{(1 - d_{i,t-1})\psi_{i,t-1}\right\}\right])$$

where $\phi_{d_{i,t-1}}$ is the probability of survival from occasion $t-1$ to t conditional on disease status on occasion $t-1$ for individual i, $\psi_{i,t}$ is the disease transmission probability, that is the probability that an individual that is uninfected on occasion $t-1$ becomes infected by occasion t, γ_t is the recruitment probability that a "pseudo-individual" is first recruited, and hence is first available for capture, on occasion t and $\alpha_{i,t}$ is a latent indicator variable of whether an individual is available to be recruited or not on occasion t. We define $\alpha_{i,t} = \left(1 - \text{I}\left(\sum_{t=1}^{t-1}(z_{i,t}) > 0\right)\right)$ such that $\alpha_{i,t} = 1$ if individual i is available to be recruited on occasion t, $\alpha_{i,t} = 0$ otherwise to ensure an individual can only be recruited once.

We note that, clearly, only individuals that are alive and uninfected can become infected and, as is the case in our motivating data, once infected, individuals cannot become uninfected. To investigate the relationship between density and disease transmission, we model $\psi_{i,t}$ as a function of population density on each occasion. We discretize the study space using a grid and create R non-overlapping habitat cells. We denote the cell in which individual AC i falls by c_i, with $c_i \in \{1, \ldots, R\}$. Local density of grid cell r, $r = 1, \ldots, R$, on occasion t is defined as $\ell_{r,t} = \sum_{i=1}^{M} I(z_{i,t} = 1, c_i = r)$, where $I(z_{i,t} = 1, c_i = r)$ is an indicator variable equal to 1 if individual i is alive on occasion t and its AC falls within cell r, and 0 otherwise. Thus, we build a logistic regression model for the probability of disease transmission

$$\text{logit}(\psi_{i,t}) = \beta_0 + \beta_1 \ell_{c_i,t} \tag{1}$$

such that an uninfected individual can become infected due to its local density ($\ell_{c_i,t}$) at the individual AC location in its habitat cell. The coefficient, β_1, determines the direction and size of the effect of local density on disease transition probability.

Finally, population size on occasion t, N_t, can be estimated as $N_t = \sum_i z_{i,t}$ and population density on occasion t, D_t, as $D_t = N_t/\text{area}(S)$, while the corresponding sizes of the infected population (N_t^{i}) and of the uninfected population (N_t^{u}) can be estimated as $N_t^{\text{i}} = \sum_i z_{i,t} d_{i,t}$ and $N_t^{\text{u}} = \sum_i z_{i,t}(1 - d_{i,t})$, respectively. Hence, disease prevalence (D_t^{i}) can be estimated as $D_t^{\text{i}} = N_t^{\text{i}}/N_t$. Density estimates for infected/uninfected individuals can be easily computed as well as realized disease density maps.

We fit models in a Bayesian framework using MCMC methods via R package NIMBLE (de Valpine et al. 2017) version 0.13.0. Additionally, to increase the computational efficiency of using a Bayesian implementation via DA, we vectorize computation and perform block sampling on correlation parameters (Turek et al. 2021) when appropriate. We employ user-defined NIMBLE functions to reduce the total number of nodes in the model and improve MCMC efficiency. We use the R package nimbleSCR (Bischof et al. 2020) version 0.2.1 to create habitat grids and for the computation of local density. To improve convergence and mixing, we use a coarse habitat grid to provide the model with a large number of latent density points to serve as a covariate on disease transmission probability. We also center latent density to improve computation by reducing the correlation between the intercept and fixed effect. Random walk block samplers are assigned to ($q_{11}^{\text{DPP}^{\circledR}}, q_{11}^{\text{Ifn}}, q_{11}^{\text{Cul}}$), ($q_{00}^{\text{DPP}^{\circledR}}, q_{00}^{\text{Ifn}}, q_{00}^{\text{Cul}}$) and ($p_{0d_i}, \sigma_{d_i}$) to improve MCMC efficiency. The code is freely available on https://github.com/xxxx/Modelling-disease-dynamics-from-spatially-explicit-capture-recapture-data. The data from the Woodchester Park study can be made available on request to Animal and Plant Health Agency via Richard Delahay.

4 | SIMULATION STUDY

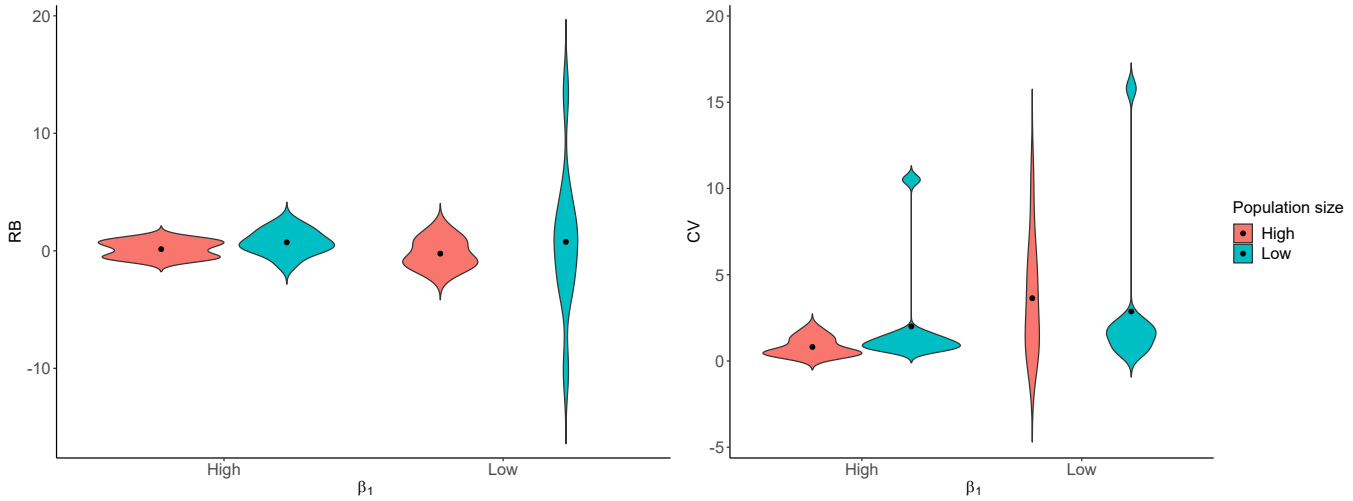
We performed a simulation study to assess the performance of the proposed modelling framework in estimating population density and all other model parameters, as well as the impact on estimation when the effect of (local) population density on transmission probability is ignored and, more importantly, when density dependent disease transmission, disease status and its effect on other model parameters are ignored altogether. We refer to our proposed model as M(ψ_ℓ), to the model that does not account for density-dependence in ψ as model M(ψ_0) and to the standard open SCR model that does not account for density dependence and disease status as model M(SRC0).

We investigate model performance at both high and low population size levels, $M = (1000, 500)$, given high and low density effects on disease transmission, $\beta_1 = (0.25, 0.1)$. We set $T = 8$ and the rest of the parameter values as: $\phi_{d_i=0} = 0.9$, $\phi_{d_i=1} = 0.8$, $\gamma_1 = 0.4$, $\gamma_2 = \gamma_{2:4} = 0.1$, $\gamma_3 = \gamma_5 = 0.2$, $\gamma_4 = \gamma_{6:T} = 0.15$, $p_{0_{d_i=0}} = 0.5$, $p_{0_{d_i=1}} = 0.2$, $\sigma_{d_i=0} = 0.5$, $\sigma_{d_i=1} = 1$, $\delta_I = 0.15$, $\beta_0 = -2.5$, $q_{11}^{\text{DPP}^{®}} = 0.492$, $q_{11}^{\text{Ifn}} = 0.809$, $q_{11}^{\text{Cul}} = 0.1$, $q_{00}^{\text{DPP}^{®}} = 0.931$, $q_{00}^{\text{Ifn}} = 0.936$ and $q_{00}^{\text{Cul}} = 0.999$. This setting results in mean($N_{1:T}$) $\approx (400, 200)$.

We used an 11×11 habitat grid in which we centered a 7×7 trapping grid resulting in a buffer width of 2 distance units. For each case, we performed 10 simulation runs and used relative bias $\left(\text{RB} = \frac{\bar{\theta}-\theta}{\theta}\right)$ to measure relative error and coefficient of variation $\left(\text{CV} = \frac{\text{SD}(\hat{\theta})}{|\bar{\theta}|}\right)$ to measure relative precision, where θ is the true parameter value, $\bar{\theta}$ is the mean and SD($\hat{\theta}$) is the standard deviation of the posterior distribution obtained, across the 10 runs.

From Tables 1 and 2 in the Supplementary Material and Figure 1 , we can see that our proposed modelling approach performed well in estimating all demographic and disease-dynamics related parameters. Estimation of the density effect (β_1) is possible (Figure 1) but is more challenging than for other parameters, as also reported by Milleret et al. (2023). As expected, the quality of inference is best in larger populations and when density effects are high. When the population size is relatively low and/or the true value of β_1 is low, estimation of β_1 is computationally challenging (Figure 1), requiring more MCMC samples to reach convergence and to obtain sufficient effective sample sizes. Consequently, this simulation study demonstrates the ability of our model to simultaneously model population demographics and disease dynamics within a spatio-temporal context.

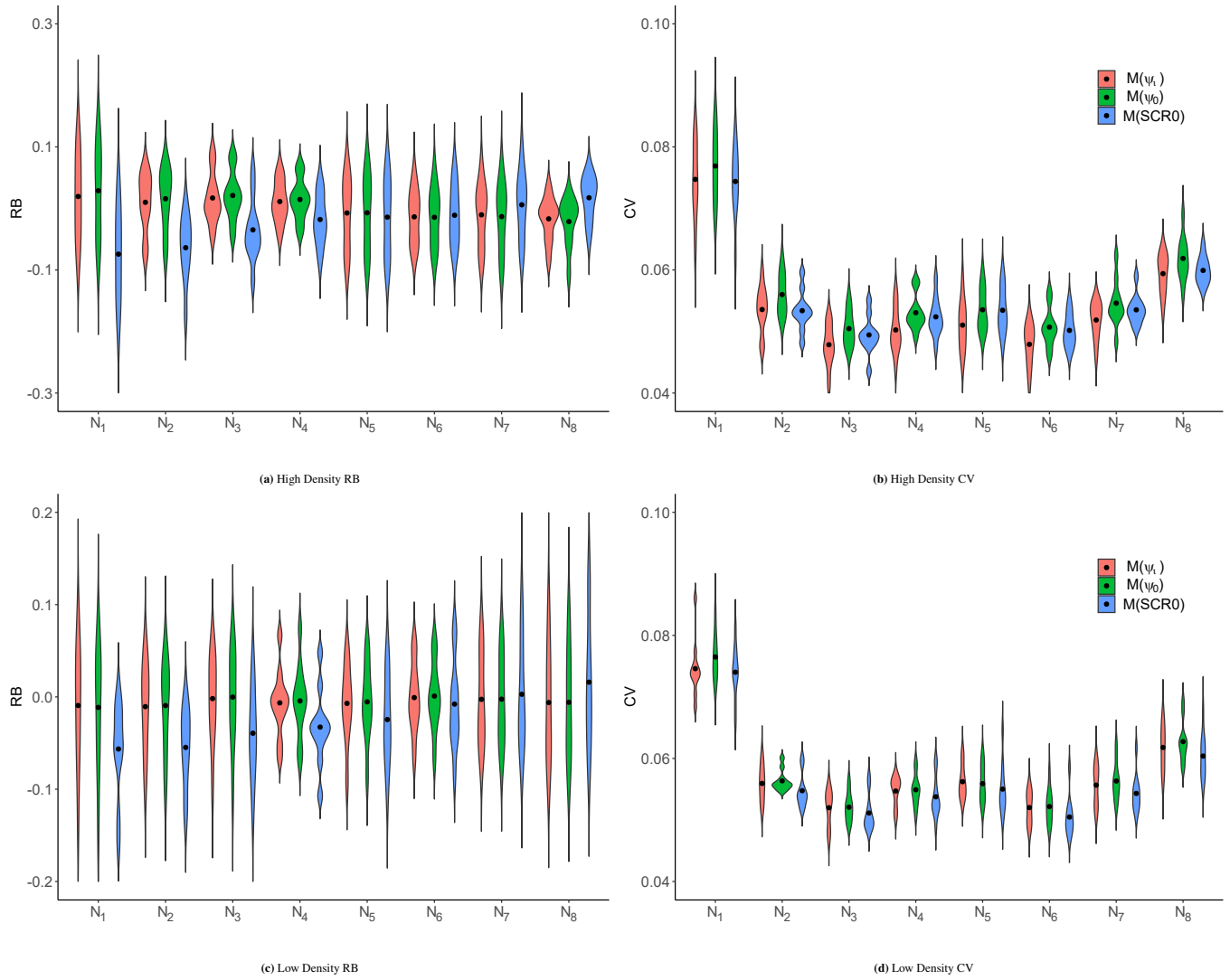
From Figures 2 and 3 , it can be seen that the M(SCR0) model underestimates population size for the first half of the study period, but tends to overestimate it at the end in all cases. This result highlights the need to account for the spatial variation in disease transmission and the disease status of individuals where there are links to demographic parameters and/or space use. All three models had similar CV for N. The M(ψ_0) model shows a small RB for N in some cases, but overall its performance is similar to the M(ψ_ℓ) model. As shown in the Supplementary Material Figure 1, we simulated the density values, which serve as the latent covariate in model M(ψ_ℓ), using a realistic scenario of small and gradual changes, with activity centres simulated from a homogeneous Poisson process, as is the standard assumption of SCR models, including the one in this paper. As a result,



(a) RB (b) CV

Figure 1 Violin plots of RB, (a), and CV, (b), for β_1 when using our proposed model, M(ψ_ℓ), given high and low population size levels. Dots represent the median in each case.

especially when the true value for β_1 is low, density does not vary dramatically between sampling occasions or between grid cells, and hence, ignoring its effect in the M(ψ_0) model does not lead to substantial bias in the estimation of N.



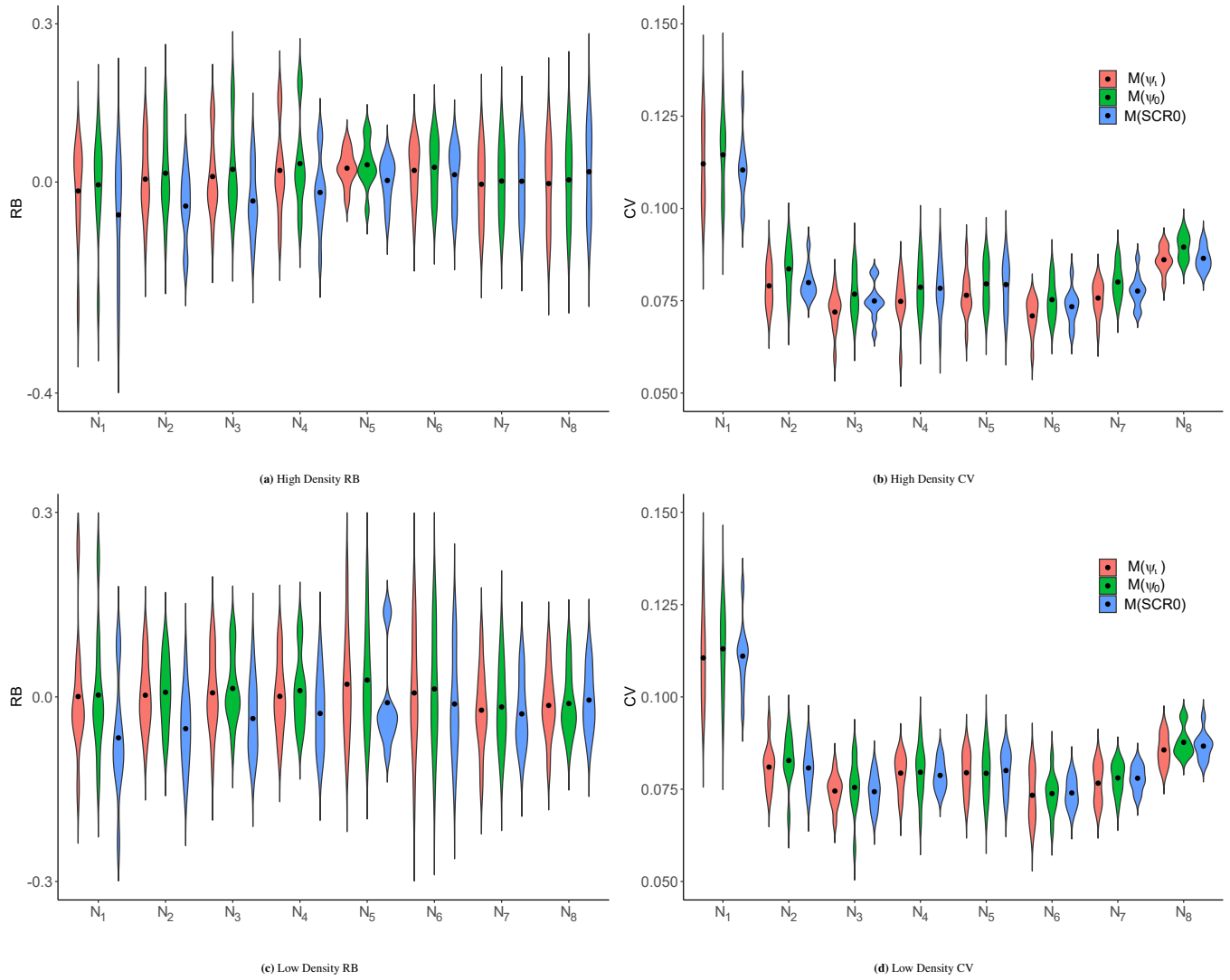
(a) High Density RB

(b) High Density CV

(c) Low Density RB

(d) Low Density CV

Figure 2 RB and CV of population size, N, at high population size data ($N \approx 400$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for three models: our proposed model, M(ψ_{ℓ}), the model that does not account for density-dependence in disease transmission, M(ψ_0) and the model that does not account for disease status, M(SCR0). Dots represent the median in each case.



(a) High Density RB

(b) High Density CV

(c) Low Density RB

(d) Low Density CV

Figure 3 RB and CV of population size, N, at low population size data ($N \approx 200$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for three models: our proposed model, M(ψ_{ℓ}), the model that does not account for density-dependence in disease transmission, M(ψ_0) and the model that does not account for disease status, M(SCR0). Dots represent the median in each case.

5 | CASE STUDY

We analyse SCR data collected from 2014 to 2018 by the long-term study at Woodchester Park using the modelling framework defined in Section 3. These years were selected as they correspond to a period when the population was undisturbed by management interventions and the diagnostic test for TB did not change, as opposed to the previous years where different tests were employed and subsequent years when badger culling for TB control was initiated in the surrounding area. We use an 11×8 habitat grid in which we place a 7×7 trapping grid. Prior settings are provided in the Supplementary Material section 2.2 where informative priors for the sensitivity and specificity of each test were elicited from Drewe et al. (2010) and Ashford et al. (2020).

Table 1 displays the most important posterior estimates obtained. Caterpillar plots of these parameter posterior summaries are also shown in Supplementary Material section 2.3. All parameters converged according to Gelman and Rubin's convergence diagnostic (Gelman & Rubin 1992), displayed good mixing, and had effective sample size (ESS) ≥ 500, with the exception of the coefficients in the model for disease transmission (ESS ≈ 100).

Table 1 Case study. Posterior summaries of model parameters.

Parameters	Mean	St.Dev	95% PCI
γ_1	0.329	0.028	(0.274, 0.384)
γ_2	0.011	0.008	(0.001, 0.031)
γ_3	0.115	0.029	(0.062, 0.173)
γ_4	0.006	0.005	(0.002, 0.020)
γ_5	0.066	0.027	(0.020, 0.123)
γ_6	0.030	0.013	(0.009, 0.057)
γ_7	0.134	0.037	(0.069, 0.217)
γ_8	0.010	0.009	(0.002, 0.035)
$\phi_{d_i=0}$	0.901	0.012	(0.877, 0.925)
$\phi_{d_i=1}$	0.811	0.029	(0.754, 0.867)
$p_{0_{d_i=0}}$	0.782	0.048	(0.693, 0.878)
$p_{0_{d_i=1}}$	0.185	0.029	(0.134, 0.248)
$\sigma_{d_i=0}$	0.267	0.008	(0.251, 0.283)
$\sigma_{d_i=1}$	0.581	0.038	(0.512, 0.664)
δ	0.210	0.043	(0.134, 0.298)
β_0	−3.712	0.573	(−5.474, −2.925)
β_1	0.065	0.195	(−0.286, 0.508)
$q_{11}^{\text{DPP®}}$	0.520	0.026	(0.471, 0.573)
q_{11}^{Ifn}	0.656	0.038	(0.587, 0.730)
q_{11}^{Cul}	0.182	0.031	(0.126, 0.243)
$q_{00}^{\text{DPP®}}$	0.978	0.007	(0.961, 0.990)
q_{00}^{Ifn}	0.907	0.013	(0.879, 0.931)
q_{00}^{Cul}	0.991	0.005	(0.978, 0.998)

Figure 4 a indicates that during the study period (2014-2018) the Woodchester badger population was in decline, with both the number of uninfected and infected individuals decreasing over the course of the study. Disease prevalence during this period remained relatively stable, albeit with some weak evidence for a decline from approximately 20% to nearer 15% (Figure 4 b). Additionally, our model output includes posterior summaries of the number of individuals newly infected on each sampling occasion (Supplementary Material Figure 17), which in this case is shown to be decreasing, in line with the decreasing population size.

Model results also confirmed known differences between infected and uninfected badgers in relation to their behaviour and survival. There was a significant difference between baseline capture and the scale parameter for uninfected and infected individuals (see Table 1), indicating that the latter were less likely to be captured at their activity center and had a larger home range area than the former.

Specifically, assuming a circular home range area, the effective home range area for an uninfected badger was estimated as 0.334 km^2 (95% PCI : (0.297 km^2, 0.377 km^2)) whilst for an infected badger was estimated as 1.586 km^2 (95% PCI : (1.234 km^2, 2.08 km^2)). Infected individuals also had a lower survival probability (ϕ) than uninfected individuals (Table 1), with the lack of overlap of the 95% PCIs indicating that these differences are statistically significant. Overall, the survival probability of infected individuals was approximately 10% lower than for uninfected individuals, albeit with more error around this estimate - potentially caused by the smaller sample size of infected individuals or greater variability in their survival. The sensitivity (q_{11}) and specificity (q_{00}) estimates for each test also provide valuable information on test performance. Specifically, Cul was found to have low sensitivity (18.2%) but had the highest specificity (99.1%), Ifn was the most sensitive (65.6%) and had relatively high specificity (90.7%), whereas the DPP® had intermediate sensitivity (52.0%) and very high specificity (97.8%). These estimates are similar to those obtained by Ashford et al. (2020) for DPP® and Drewe et al. (2010) for Ifn and Cul and reflect known differences in the performance of the tests and the biological processes being targeted (e.g. Ifn detecting an initial cellular response to infection while Cul detects bacterial shedding by infectious individuals). Hence, the tests may in some cases be identifying animals at different stages of infection.

Figure 5 displays the population density maps for infected and uninfected individuals across years in spring after cubs have been recruited to the population. These plots are standardized across years with the black dots representing the setts trapped. These outputs reveal spatio-temporal variation in the density of uninfected and infected individuals across the population. High densities of infected badgers were concentrated in the central (and northern) and western area of the study site at the start of our study period, becoming more diffuse over time. The eastern parts of the study site maintained consistently higher densities of uninfected badgers throughout the period.

Finally, density is estimated to have a weak positive effect on the probability of disease transmission that is not clearly different from zero (wide 95% PCI that includes 0). The simplistic interpretation would be that transmission was independent of local

population density during our study period. However, as discussed in Section 4, density is a latent variable with an unknown effect, and hence the power to detect small effects relies heavily on the number of sampling occasions and the number of individuals. Consequently, we can interpret this finding as evidence that strong density-dependence of transmission is highly unlikely, but that transmission could instead either be weakly density-dependent or close to frequency-dependent.

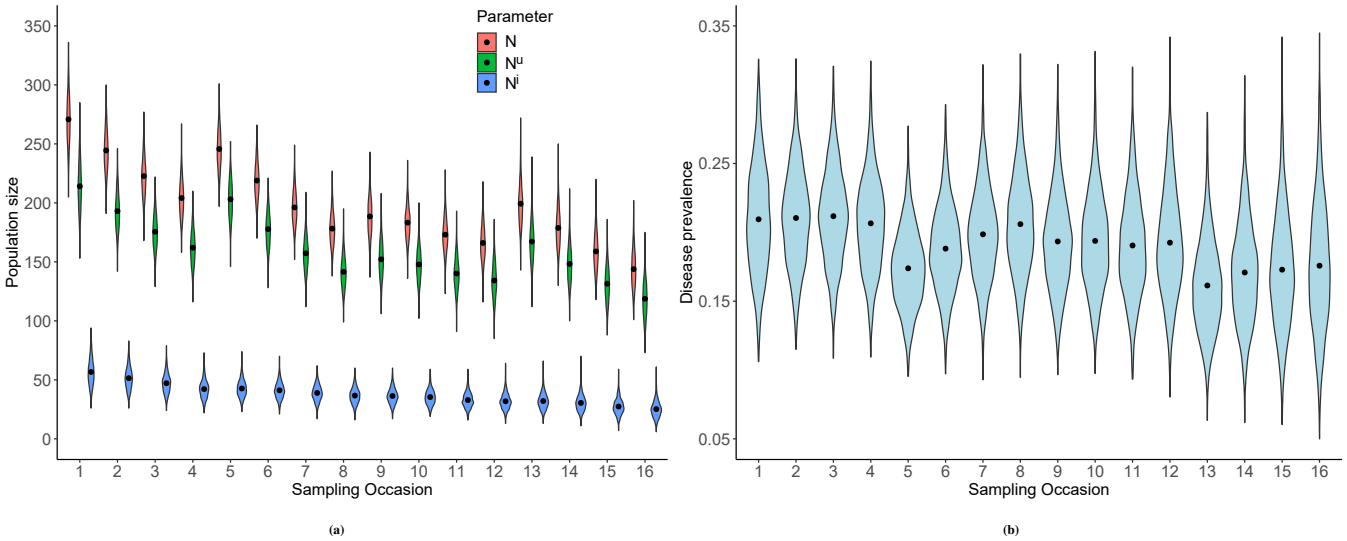
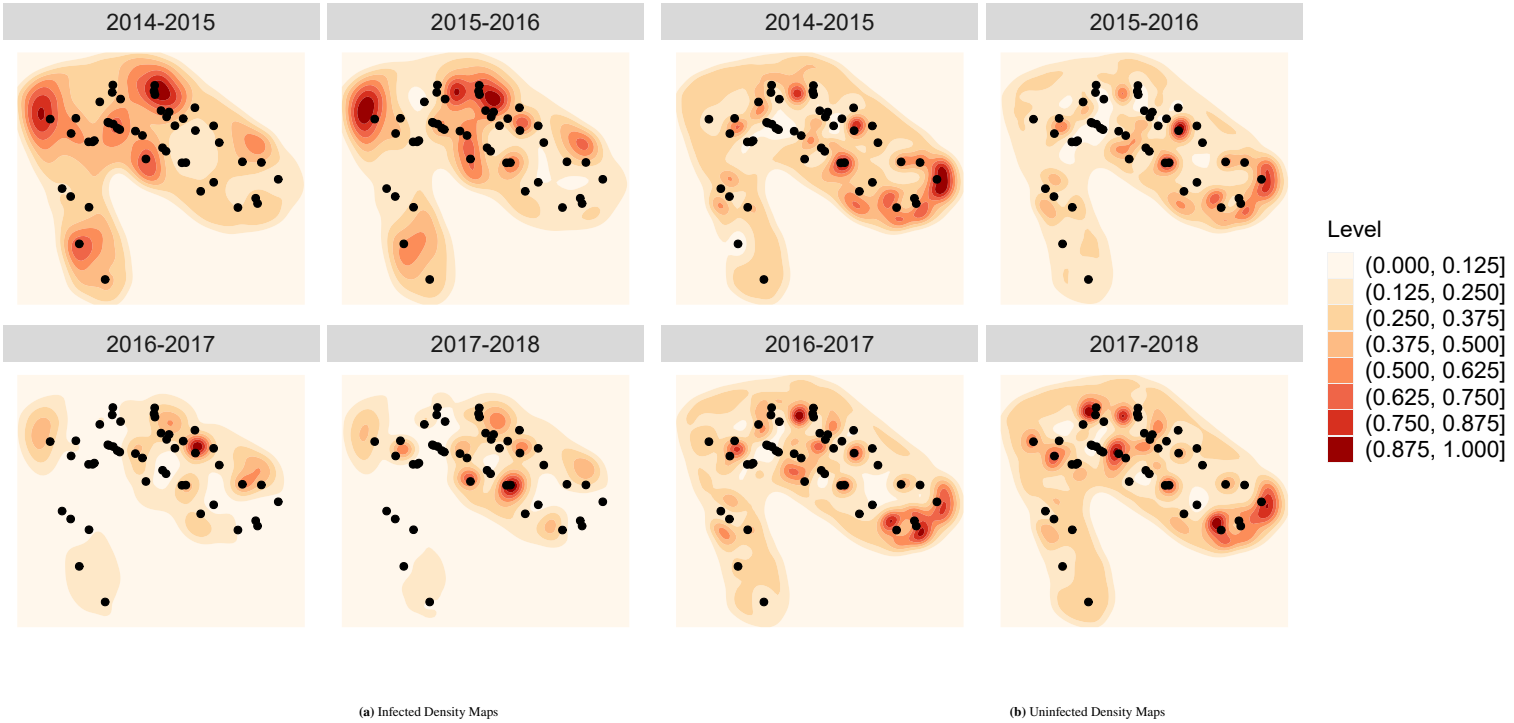


Figure 4 Violin plots, (a), of total (N_t), uninfected (N_t^u), infected (N_t^i) population size, and (b) disease prevalence. Dots represent the posterior median.

6 | DISCUSSION

We have developed a novel SCR model that uses disease data from multiple imperfect tests together with SCR data to simultaneously model population demographics and disease dynamics within a spatiotemporal context. Accounting for observation error in both the individual capture process and the disease testing process, our modelling approach accounts for spatial variation in survival and individual capture probabilities as a function of individual (latent) disease state as well as variation in disease transmission as a function of population density. This allows for a better understanding of how disease dynamics relate to population demographics in a spatiotemporal context.

We also conducted an extensive simulation study to assess model performance for a number of scenarios. Our simulations generated encouraging results for our modelling approach and highlighted that, if spatial variation in disease transmission and heterogeneity in demographic rates (capture and survival probabilities) induced by individual disease status are not accounted for, biased estimates of population density can be produced. Notably, there are existing models that use finite mixtures to model heterogeneity in these demographic rates (Pledger, Pollock, & Norris 2010). We have not considered these models but it is likely



(a) Infected Density Maps

(b) Uninfected Density Maps

Figure 5 Standardized density maps for infected, (a), and uninfected, (b), individuals across years in Spring each year. Black dots represent the setts trapped and higher level values indicate higher density and vice-versa.

that they might return similar inference on population density to our proposed model. However, such models do not provide information on density-dependent disease transmission or information on individual disease status effects on such demographic parameters. Finally, our model assumes that individual home range changes once an individual becomes infected, but without allowing for additional individual heterogeneity in home range we cannot test if for at least some individuals change in home range preceded disease.

Applying this new model to a dataset on European badgers, naturally infected with *M. bovis*, our model provided valuable insights into badger-*M. bovis* ecology with broader implications for wildlife disease ecology in general Our model results agreed closely with previous findings from the Woodchester Park study system. Estimates of population size align with those from the long-term study (R. J. Delahay et al. 2013; McDonald et al. 2018) and support the observed recent decline, whilst our estimates of individual badger home range area are also similar to those from previous studies of this population (Garnett, Delahay, and Roper (2005), Tuyttens et al. (2000)). Finally, our estimates of disease prevalence and incidence, as well as disease-associated changes in survival are similar to those found previously for this study population (prevalence: R. J. Delahay et al. (2013); incidence and changes in survival: Graham et al. (2013). However, the approach considered in this paper is the first to simultaneously model all these processes from the available SCR data and hence properly account for uncertainty in every stage of the process.

One key advantage of our modelling framework is that by using longitudinal diagnostic test results to infer the (unobservable) disease status of an individual, we are able to gain new insights into how individual movement patterns vary with disease status.

Specifically, our results show that infected badgers ranged over larger areas than uninfected badgers. Previous research has detected a tendency for test-positive individuals (i.e. those likely to be infected) to make greater use of outlying setts (Weber, Bearhop, et al. 2013) and have more between-group contacts (Silk et al. 2018; Weber, Carter, et al. 2013), both of which are traits expected to be linked to wider ranging behaviour. Our model results indicate these changes may (to some extent) be linked directly to infection. The tendency for infected badgers to start ranging further likely has important implications for the epidemiology of the badger-*M. bovis* host-pathogen system. Due to the modular nature of badger contact networks (Rozins et al. 2018), movements between groups offer important opportunities for transmission that enable wider pathogen spread. Therefore, changes in the behaviour of infected badgers could play an important role in the longer-term persistence of the disease. Previous research in the Woodchester system has revealed a positive association between new individuals arriving in a group and the incidence of disease (Vicente, Delahay, Walker, & Cheeseman 2007). Infected badgers ranging over larger areas than uninfected badgers, using areas that are larger than typical social group territory size, would provide a mechanism to explain these findings; it is disproportionately likely that an individual moving between groups is infected relative to the overall population.

Our results also add to our knowledge of how transmission scales with population density in the badger-*M. bovis* system. We find no clear relationship between local population density and the incidence of infection in the Woodchester Park badger population (R. J. Delahay, Smith, Ward, & Cheeseman 2005; R. J. Delahay et al. 2013) . Historically, many infections that spread by (non-sexual) close contact were assumed to exhibit density-dependent (as opposed to frequency-dependent) transmission, and this principle of density-dependence underpins some interventions to control disease in wildlife populations (Carter et al. 2009; McCallum 2016). More recently, studies have more commonly considered a continuum between frequency-dependent and density-dependent transmission driven by changes in individual behaviour (Hopkins et al. 2020). For example, our result of lack of evidence for density-dependence transmission suggested that it may be best to consider the transmission of many infectious diseases including tuberculosis to be a function of population density at low population densities (i.e. density-dependent) but independent of it (i.e. frequency-dependent) at high population densities (Hu, Nigmatulina, & Eckhoff 2013). Our results fit well with this general pattern given that the Woodchester Park badger population was relatively high density during the study period, compared with badger populations elsewhere. At this population scale, it is likely that the social structure of the population plays an important role and it would be valuable for future research to focus on this question at finer social and spatial scales.

Finally, by separately estimating the density of uninfected and infected badgers in the population, our SCR model provides an intuitive approach to the analysis of spatiotemporal variation in *M. bovis* epidemiology in the population whilst accounting for uncertainty generated from the use of capture-recapture data (i.e. imperfect capture of badgers) and the limitations of diagnostic test data (i.e. imperfect knowledge of disease state). The local density maps generated (Figure 5) can provide a useful tool for visualizing hotspots of disease (i.e. areas with a high prevalence of infected individuals) or guiding surveillance (e.g. by

revealing areas with rapidly increasing or decreasing prevalence). Consequently, our study highlights the value of integrating disease status within an SCR framework for applied disease ecology more generally.

The choice of grid size is a crucial factor in this modelling framework as it can impact accuracy and computational efficiency. A smaller grid size provides finer resolution, capturing intricate details and small-scale patterns in density. However, using a smaller grid size comes at the cost of increased computational complexity and memory requirements. On the other hand, a larger grid size provides a coarser resolution that can overlook smaller-scale patterns but reduces the computational burden. Thus, it is important to strike a balance between accuracy and computational efficiency when selecting grid size. To achieve this, a sensitivity analysis can be carried out. Varying the grid size helps determine the most appropriate grid size and also helps ensure that the chosen grid size does not unduly influence the results. At the same time, since disease transmission is dependent on latent density, the grid size needs to be chosen such that there are adequate latent density points to serve as a covariate on disease transmission.

One caveat to our model is that we assumed individual activity centres do not change over the period of study. This reflected limited badger movement as highlighted by Rogers et al. (1998) and in the Supplementary Material Figure 12. However, this assumption will be violated for species that change activity centres frequently. Our modelling approach can be extended to accommodate such movement by using different state models such as the independent (J. Royle, Chandler, Sollmann, & Gardner 2014) and the Markovian random walk model (Raabe, Gardner, & Hightower 2014). We have also assumed independence between individuals in terms of the locations of their activity centres, however, populations can exhibit attraction or repulsion, which means that the spatial pattern cannot be described by a homogeneous Poisson process. In these cases, our model would need to be extended using models that account for repulsion (Diana, Matechou, Griffin, Jhala, & Qureshi 2022) and/or attraction (McLaughlin & Bar 2021).

Another avenue for future work is the modelling of disease transmission. Here we assumed that once infected an individual remained so for the rest of its life, consistent with the normal approach when modelling *M. bovis* transmission in badgers. However, in other cases it may be important to introduce further states to the disease model, such as a recovered state (to represent individuals in which the disease has resolved and individuals may have acquired protection from re-infection) or a vaccinated state. Such multistate disease models have been fitted to capture-recapture data (e.g Marescot et al. 2018), and could be easily incorporated within our modelling framework.

It could also be possible to vary how disease transmission probability is associated with the spatiotemporal distribution of infected and uninfected individuals. In this paper, we used a logistic regression to investigate the effect of (latent) local density on disease transmission probability and induce heterogeneity. However, other latent variables of this type could also be considered. For example, one might consider alternative models, for example a model which assumes that an uninfected individual is more likely to become infected the closer it is to infected individuals, or a model which considers disease transmission probability

as a function of the overlap of home range areas between an uninfected individual and surrounding infected individuals. Thus, future work can be focused on developing such models and comparing model performance.

In conclusion, our SCR model provides a novel tool to investigate the relationship between population demographics, spatial behaviour, and infectious disease dynamics in imperfectly sampled systems. By applying it to other host-pathogen systems, it may be possible to gain valuable insights into how spatial behaviour and pathogen epidemiology are interwoven with important implications for wildlife disease ecology and management.

References

Ashford, R. T., Anderson, P., Waring, L., Davé, D., Smith, F., Delahay, R. J., … Lesellier, S. (2020). Evaluation of the Dual Path Platform (DPP) VetTB assay for the detection of Mycobacterium bovis infection in badgers. *Preventive veterinary medicine*, *180*, 105005.

Bischof, R., Turek, D., Milleret, C., Ergon, T., Dupont, P., & de Valpine, P. (2020). nimbleSCR: spatial capture-recapture (SCR) methods using 'nimble'. *R package version 0.1. 0*.

Borchers, D. L., & Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics*, *64*(2), 377–385.

Buzdugan, S. N., Vergne, T., Grosbois, V., Delahay, R. J., & Drewe, J. A. (2017). Inference of the infection status of individuals using longitudinal testing data from cryptic populations: Towards a probabilistic approach to diagnosis. *Scientific Reports*, *7*(1), 1–11.

Carter, S. P., Roy, S. S., Cowan, D. P., Massei, G., Smith, G. C., Ji, W., … Delahay, R. J. (2009). Options for the control of disease 2: targeting hosts. *Management of disease in wild mammals*, 121–146.

Cheeseman, C., CL, C., et al. (1982). Methods of marking badgers (Meles meles).

Choquet, R., Carrié, C., Chambert, T., & Boulinier, T. (2013). Estimating transitions between states using measurements with imperfect detection: application to serological data. *Ecology*, *94*(10), 2160–2165.

Dalley, D., Davé, D., Lesellier, S., Palmer, S., Crawshaw, T., Hewinson, R. G., & Chambers, M. (2008). Development and evaluation of a gamma-interferon assay for tuberculosis in badgers (Meles meles). *Tuberculosis*, *88*(3), 235–243.

Delahay, R., Carter, S., Forrester, G., Mitchell, A., & Cheeseman, C. (2006). Habitat correlates of group size, bodyweight and reproductive performance in a high-density Eurasian badger (Meles meles) population. *Journal of Zoology*, *270*(3), 437–447.

Delahay, R. J., Smith, G. C., Ward, A. I., & Cheeseman, C. L. (2005). Options for the management of bovine tuberculosis transmission from badgers (Meles meles) to cattle: evidence from a long-term study. *Mammal Study*, *30*(Supplement), S73–S81.

Delahay, R. J., Walker, N., Smith, G., Wilkinson, D., Clifton-Hadley, R., Cheeseman, C., … Chambers, M. (2013). Long-term temporal trends and estimated transmission rates for Mycobacterium bovis infection in an undisturbed high-density badger (Meles meles) population. *Epidemiology & Infection*, *141*(7), 1445–1456.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*(2), 403–413.

Diana, A., Matechou, E., Griffin, J. E., Jhala, Y., & Qureshi, Q. (2022). A vector of point processes for modeling interactions between and within species using capture-recapture data. *Environmetrics*, e2781.

Drewe, J. A., Tomlinson, A. J., Walker, N. J., & Delahay, R. J. (2010). Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. *PloS one*, *5*(6), e11196.

Efford, M. (2004). Density estimation in live-trapping studies. *Oikos*, *106*(3), 598–610.

Enøe, C., Georgiadis, M. P., & Johnson, W. O. (2000). Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive veterinary medicine*, *45*(1-2), 61–81.

Gallagher, J., & Horwill, D. (1977). A selective oleic acid albumin agar medium for the cultivation of Mycobacterium bovis. *Epidemiology & Infection*, *79*(1), 155–160.

Garnett, B., Delahay, R., & Roper, T. (2005). Ranging behaviour of European badgers (Meles meles) in relation to bovine tuberculosis (Mycobacterium bovis) infection. *Applied Animal Behaviour Science*, *94*(3-4), 331–340.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.

Graham, J., Smith, G., Delahay, R., Bailey, T., McDonald, R., & Hodgson, D. (2013). Multi-state modelling reveals sex-dependent transmission, progression and severity of tuberculosis in wild badgers. *Epidemiology & Infection*, *141*(7), 1429–1436.

Hopkins, S. R., Fleming-Davies, A. E., Belden, L. K., & Wojdak, J. M. (2020). Systematic review of modelling assumptions and empirical evidence: does parasite transmission increase nonlinearly with host density? *Methods in Ecology and Evolution*, *11*(4), 476–486.

Hu, H., Nigmatulina, K., & Eckhoff, P. (2013). The scaling of contact rates with population density for the infectious disease models. *Mathematical biosciences*, *244*(2), 125–134.

Krkošek, M. (2010). Host density thresholds and disease control for fisheries and aquaculture. *Aquaculture Environment Interactions*, *1*(1), 21–32.

Manlove, K., Cassirer, E. F., Cross, P. C., Plowright, R. K., & Hudson, P. J. (2016). Disease introduction is associated with a phase transition in bighorn sheep demographics. *Ecology*, *97*(10), 2593–2602.

Marescot, L., Benhaiem, S., Gimenez, O., Hofer, H., Lebreton, J.-D., Olarte-Castillo, X. A., … East, M. L. (2018). Social status mediates the fitness costs of infection with canine distemper virus in Serengeti spotted hyenas. *Functional Ecology*, *32*(5), 1237–1250.

McCallum, H. (2016). Models for managing wildlife disease. *Parasitology*, *143*(7), 805–820.

McCallum, H., Tompkins, D. M., Jones, M., Lachish, S., Marvanek, S., Lazenby, B., … Hawkins, C. E. (2007). Distribution and impacts of Tasmanian devil facial tumor disease. *EcoHealth*, *4*, 318–325.

McDonald, J. L., Robertson, A., & Silk, M. J. (2018). Wildlife disease ecology from the individual to the population: Insights

from a long-term study of a naturally infected European badger population. *Journal of Animal Ecology*, *87*(1), 101–112.

McLaughlin, P., & Bar, H. (2021). A spatial capture–recapture model with attractions between individuals. *Environmetrics*, *32*(1), e2653.

Milleret, C., Dey, S., Dupont, P., Brøseth, H., Turek, D., de Valpine, P., & Bischof, R. (2023). Estimating spatially variable and density-dependent survival using open-population spatial capture–recapture models. *Ecology*, *104*(2), e3934.

O'Neill, X., White, A., Gortázar, C., & Ruiz-Fons, F. (2023). The Impact of Host Abundance on the Epidemiology of Tick-Borne Infection. *Bulletin of Mathematical Biology*, *85*(4), 30.

Pledger, S., Pollock, K. H., & Norris, J. L. (2010). Open capture–recapture models with heterogeneity: II. Jolly–Seber model. *Biometrics*, *66*(3), 883–890.

Raabe, J. K., Gardner, B., & Hightower, J. E. (2014). A spatial capture–recapture model to estimate fish survival and location from linear continuous monitoring arrays. *Canadian Journal of Fisheries and Aquatic Sciences*, *71*(1), 120–130.

Rogers, L., Cheeseman, C., Mallinson, P., & Clifton-Hadley, R. (1997). The demography of a high-density badger (Meles meles) population in the west of England. *Journal of Zoology*, *242*(4), 705–728.

Rogers, L., Delahay, R., Cheeseman, C., Langton, S., Smith, G., & Clifton-Hadley, R. (1998). Movement of badgers (Meles meles) in a high–density population: individual, population and disease effects. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1403), 1269–1276.

Royle, J., Chandler, R., Sollmann, R., & Gardner, B. (2014). *Spatial capture-recapture. Waltham, Massachusetts.* Elsevier, Academic Press.

Royle, J. A., & Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *Journal of Ornithology*, *152*(2), 521–537.

Royle, J. A., Fuller, A. K., & Sutherland, C. (2018). Unifying population and landscape ecology with spatial capture–recapture. *Ecography*, *41*(3), 444–456.

Royle, J. A., & Young, K. V. (2008). A hierarchical model for spatial capture–recapture data. *Ecology*, *89*(8), 2281–2289.

Rozins, C., Silk, M. J., Croft, D. P., Delahay, R. J., Hodgson, D. J., McDonald, R. A., … Boots, M. (2018). Social structure contains epidemics and regulates individual roles in disease transmission in a group-living mammal. *Ecology and Evolution*, *8*(23), 12044–12055.

Silk, M. J., & Fefferman, N. H. (2021). The role of social structure and dynamics in the maintenance of endemic disease. *Behavioral Ecology and Sociobiology*, *75*(8), 122.

Silk, M. J., Hodgson, D. J., Rozins, C., Croft, D. P., Delahay, R. J., Boots, M., & McDonald, R. A. (2019). Integrating social behaviour, demography and disease dynamics in network models: applications to disease management in declining wildlife populations. *Philosophical Transactions of the Royal Society B*, *374*(1781), 20180211.

Silk, M. J., Weber, N. L., Steward, L. C., Hodgson, D. J., Boots, M., Croft, D. P., … McDonald, R. A. (2018). Contact networks structured by sex underpin sex-specific epidemiology of infection. *Ecology letters*, *21*(2), 309–318.

Sutherland, C., Royle, J. A., & Linden, D. W. (2019). oSCR: A Spatial Capture-Recapture R Package for Inference about Spatial Ecological Processes. *Ecography*.

Turek, D., Milleret, C., Ergon, T., Brøseth, H., Dupont, P., Bischof, R., & De Valpine, P. (2021). Efficient estimation of large-scale spatial capture–recapture models. *Ecosphere*, *12*(2), e03385.

Tuyttens, F., Delahay, R., Macdonald, D., Cheeseman, C., Long, B., & Donnelly, C. (2000). Spatial perturbation caused by a badger (Meles meles) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (Mycobacterium bovis). *Journal of Animal Ecology*, *69*(5), 815–828.

Vicente, J., Delahay, R., Walker, N., & Cheeseman, C. (2007). Social organization and movement influence the incidence of bovine tuberculosis in an undisturbed high-density badger Meles meles population. *Journal of Animal Ecology*, *76*(2), 348–360.

Vredenburg, V. T., Knapp, R. A., Tunstall, T. S., & Briggs, C. J. (2010). Dynamics of an emerging disease drive large-scale amphibian population extinctions. *Proceedings of the National Academy of Sciences*, *107*(21), 9689–9694.

Weber, N., Bearhop, S., Dall, S. R., Delahay, R. J., McDonald, R. A., & Carter, S. P. (2013). Denning behaviour of the European badger (Meles meles) correlates with bovine tuberculosis infection status. *Behavioral Ecology and Sociobiology*, *67*(3), 471–479.

Weber, N., Carter, S. P., Dall, S. R., Delahay, R. J., McDonald, J. L., Bearhop, S., & McDonald, R. A. (2013). Badger social networks correlate with tuberculosis infection. *Current Biology*, *23*(20), R915–R916.

Woodroffe, R., Donnelly, C. A., Cox, D., Bourne, F. J., Cheeseman, C., Delahay, R., … Morrison, W. I. (2006). Effects of culling on badger Meles meles spatial organization: implications for the control of bovine tuberculosis. *Journal of Applied Ecology*, *43*(1), 1–10.