That's not the Mona Lisa! How to interpret spatial capture-recapture density surface estimates

David L. Borchers[1,*], Ian Durbach[1], Rishika Chopara[2], Ben C. Stevenson[2], Rachel Phillip[1], and Koustubh Sharma[3]

[1]Centre for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, Univeristy of St Andrews, The Observatory, St Andrews, Fife, KY16 9LZ, Scotland

[2]Department of Statistics, University of Auckland, Auckland 1010, New Zealand

[3]Snow Leopard Trust, Seattle, Washington, United States of America

[*]Corresponding author: dlb@st-andrews.ac.uk

S1 Simulation study for Bayesian models

Results presented in Section 4 were generated by fitting maximum-likelihood SCR models to simulated data. In this appendix we reproduce results from Section 4 using Bayesian models fitted via MCMC to demonstrate that our results are not simply a consequence of adopting a classical approach. In Section S1.1 we describe our Baysian models, in Section S1.2 we present results of our simulation study, and in Section S1.3 we discuss similarities and differences between these results and those presented in the main manuscript based on maximum-likelihood models.

S1.1 Model fitting

We fitted Bayesian versions of the maximum-likelihood models presented in Section 4 to each data set. Again, we used models with constant density to estimate realised AC and realised usage surfaces, and a model with inhomogeneous density characterised by a log-linear relationship with a spatial covariate to estimate expected AC density surfaces.

We fitted our models in NIMBLE (insert reference) using data augmentation (Tanner & Wong, 1987), which has become the prevailing way to fit SCR models under a Bayesian framework. This approach involves sampling a superpopulation of M activity centres, including those of the n animals detected on the SCR survey. We have an indicator variable z_i for the ith animal, denoting whether the ith animal in the augmented population is 'exists' in a given MCMC iteration. Rather than directly estimating N, the population size, we estimate the data augmentation parameter, ψ, the expected proportion of the animals in the superpopulation for which the indicator is equal to 1. For each MCMC iteration we obtain a sample from the posterior of N using $\sum_{\imath=1}^{M} z_i$. A sample from the posterior for animal density can be obtained by dividing by the area of the survey region. Further details on data augmentation can be found in Kéry & Schaub (2012, pp. 139–157).

Rishika, please check this is the correct reference.

We used the following uninformative priors for the detection function parameters, specifying a prior for $\log\{1/(2\sigma^2)\}$ rather than σ directly:

$$\lambda_0 \sim \text{Gamma}(0.001, 0.001)$$
$$\log\left(\frac{1}{2\sigma^2}\right) \sim \text{Uniform}(-10, 10)$$

For the constant density model, the activity centres were given a uniform prior distribution over the survey region and the data augmentation parameter

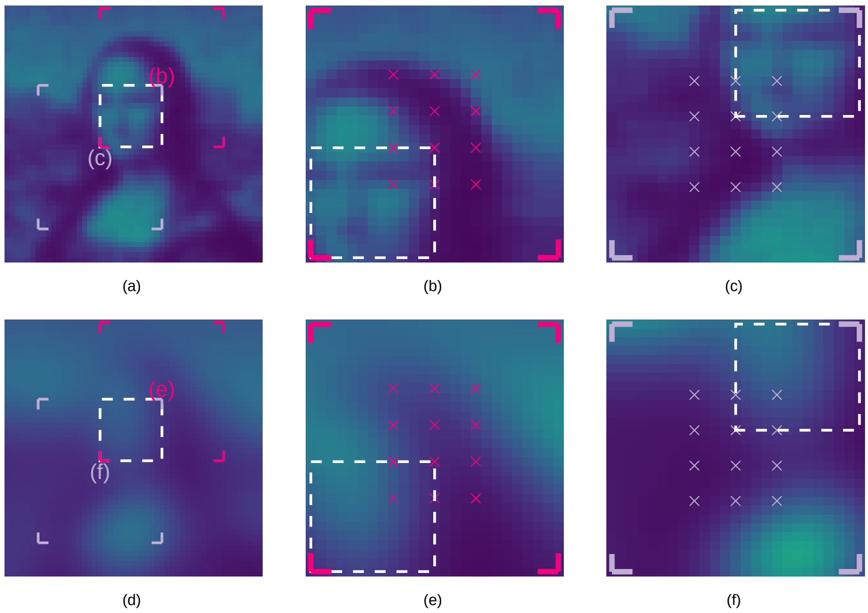

Figure 1: Figure 7

was given a uniform prior from 0 to 1. For the inhomogeneous density model, animal density at location $\boldsymbol{x}$ is given by $D(\boldsymbol{x}) = \exp\{\beta_0 + \beta_1 y(\boldsymbol{x})\}$, where $y(\boldsymbol{x})$ is a measurement of a covariate at location $\boldsymbol{x}$. We used the following uninformative priors for the coefficients β_0 and β_1:

$$\beta_0 \sim \text{Uniform}(-10, 10)$$
$$\beta_1 \sim \text{Uniform}(-10, 10)$$

When we fit each constant density model, we ran () MCMC iterations, where we set M to be equal to (). We also used a thinning value of (), an adapatation interval of () and () burn-in iterations.

When fitting each inhomogeneous density model, we ran () MCMC iterations, and used a value of () for M. The thinning value was (), with an adaptation interval of () and () burn-in iterations.

S1.2 Results

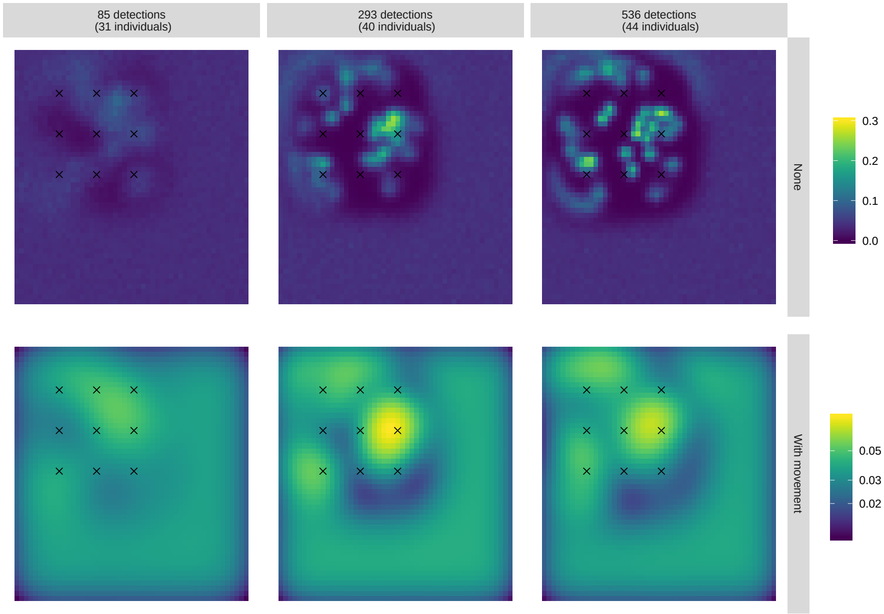

Figure 2: Figure 9

S1.3 Discussion

S2 Estimation of realised usage density

Estimation of realised usage density is a similar process for both maximum likelihood and Bayesian approaches: we sum usage densities for each individual animal, each of which is calculated by convolving the estimated PDF of its activity centre with an individual usage distribution.

S2.1 The maximum likelihood approach

For maximum likelihood, the estimated usage density for the ith animal, with capture history $\boldsymbol{\omega}_i$, is given by

$$f_{\boldsymbol{s}|\boldsymbol{\omega}}(\boldsymbol{s} \mid \boldsymbol{\omega}_i; \widehat{\boldsymbol{\theta}}) = \int f_{\boldsymbol{x}|\boldsymbol{\omega}}(\boldsymbol{x} \mid \boldsymbol{\omega}_i; \widehat{\boldsymbol{\theta}}) f_{\boldsymbol{s}|\boldsymbol{x}}(\boldsymbol{s} \mid \boldsymbol{x}; \widehat{\boldsymbol{\theta}}) d\boldsymbol{x}, \tag{1}$$

where

- $\widehat{\boldsymbol{\theta}}$ is a vector containing the maximum likelihood estimates of the encounter function parameters;
- $f_{\boldsymbol{s}|\boldsymbol{\omega}}(\boldsymbol{s} \mid \boldsymbol{\omega}_i; \widehat{\boldsymbol{\theta}})$ is the estimated usage distribution, providing the probability density of finding an individual with capture history $\boldsymbol{\omega}_i$ at location $\boldsymbol{s}$ at a randomly selected point in time;
- $f_{\boldsymbol{x}|\boldsymbol{\omega}}(\boldsymbol{x} \mid \boldsymbol{\omega}_i; \widehat{\boldsymbol{\theta}})$ is the estimated PDF of the activity centre of an individual with capture history $\boldsymbol{\omega}_i$ (see Section 3); and
- $f_{\boldsymbol{s}|\boldsymbol{x}}(\boldsymbol{s} \mid \boldsymbol{x}; \widehat{\boldsymbol{\theta}})$ is the estimated usage distribution of the individual conditional on the activity centre, providing the probability density of the individual being at location $\boldsymbol{s}$ given that its activity centre is at $\boldsymbol{x}$.

Estimated usage density at location $\boldsymbol{s}$ is then given by $\widehat{D}_u(\boldsymbol{s}) = \sum_i f_{\boldsymbol{s}|\boldsymbol{\omega}}(\boldsymbol{s} \mid \boldsymbol{\omega}_i; \widehat{\boldsymbol{\theta}})$, noting that the sum is over individuals that were not detected, with capture histories $(0, \cdots, 0)$, along with those that were.

It's unclear to me whether we directly estimate a detection function or an encounter function in our models. Below I assume the reader will know what an encounter fucntion is, but it might need to be explained more explicitly. It's important that we construct the individual usage distribution using an encounter function rather than a detection function, because the rate at which an animal visits a location is proportional to the encounter function, but not to the detection function.

Here we constructed the individual usage distribution under the assumption that the density of an individual being at location $\boldsymbol{s}$ given its activity centre is at $\boldsymbol{x}$ is proportional to the encounter function $h\{d(\boldsymbol{s}, \boldsymbol{x}); \widehat{\boldsymbol{\theta}}\}$, where $d(\boldsymbol{s}, \boldsymbol{x})$ is the Euclidean distance between $\boldsymbol{s}$ and $\boldsymbol{x}$, and so

$$f_{\boldsymbol{s}|\boldsymbol{x}}(\boldsymbol{s} \mid \boldsymbol{x}; \widehat{\boldsymbol{\theta}}) = \frac{h\{d(\boldsymbol{s}, \boldsymbol{x}); \widehat{\boldsymbol{\theta}}\}}{\int h\{d(\boldsymbol{s}', \boldsymbol{x}); \widehat{\boldsymbol{\theta}}\} d\boldsymbol{s}'}, \tag{2}$$

where the denominator is a normalising constant.

S2.2 The Bayesian approach

Bayesian models fitted via MCMC can directly sample activity centres of detected individuals, and also of undetected individuals using data augmentation, thus obtaining samples from $f_{\boldsymbol{x}|\boldsymbol{\omega}}(\boldsymbol{x} \mid \boldsymbol{\omega})$ for each individual. We can use these samples directly to obtain the following approximation of the ith individual's usage distribution:

$$f_{\boldsymbol{s}|\boldsymbol{\omega}}(\boldsymbol{s} \mid \boldsymbol{\omega}_i) \approx \frac{1}{J} \sum_{j=1}^{J} f_{\boldsymbol{s}|\boldsymbol{x}}(\boldsymbol{s} \mid \boldsymbol{x}_{(j)}, \boldsymbol{\theta}_{(j)}), \tag{3}$$

where $\boldsymbol{x}_{(j)}$ and $\boldsymbol{\theta}_{(j)}$ are the activity centre and a vector of encounter function parameters that were sampled on the jth of J total MCMC iterations, respectively. The estimated usage distribution is therefore not conditional on one particular set of estimated parameter values, but instead considers the range of values across the posterior distribution of $\boldsymbol{\theta}$.

S2.3 Discussion

I'm not sure that this is the best place for the discussion below, but leaving it here for now.

We constructed individual usage distributions using the encounter function from our SCR model, but this may not always be appropriate. For example, if individuals cannot fully explore their home range within the duration of the survey, then we would not expect the spatial range of the detection function to match the extent of an animal's usage distribution.

Even for longer surveys, it may not be sensible to relate the range of the encounter function to the size of the region used by an individual even for longer surveys, so care should be taken when this practice is used. For example, Tenan, Pedrini, Bragalanti, Groff & Sutherland (2017) found that the spatial scale of the encounter rate function for brown bears (*Ursus arctos*) estimated using SCR was not consistent with spatial usage parameters estimated from other data sources, although Popescu, de Valpine & Sweitzer (2014) did not detect any such inconsistency for a population of fishers (*Pekania pennanti*). If alternative data sources are available (e.g., telemetry, or opportunistic data such as hair or scat samples) they may be incoprorated for improved estimation of individual usage distributions (Tenan *et al.*, 2017).

Our method also assumes that home ranges are circular, however their shapes are likely to be modified by variables relating to population and landscape connectivity (see Drake, Lambin & Sutherland, in press, for a review).

References

Drake, J., Lambin, X. & Sutherland, C. (in press) The value of considering demographic contributions to connectivity: a review. *Ecography.*

Kéry, M. & Schaub, M. (2012) *Bayesian Population Analysis using WinBUGS.* Academic Press, Oxford.

Popescu, V.D., de Valpine, P. & Sweitzer, R.A. (2014) Testing the consistency of wildlife data types before combining them: the case of camera traps and telemetry. *Ecology and Evolution*, **4**, 933–943.

Tanner, M.A. & Wong, W.H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528–540.

Tenan, S., Pedrini, P., Bragalanti, N., Groff, C. & Sutherland, C. (2017) Data integration for inference about spatial processes: A model-based approach to test and account for data inconsistency. *PlosOne*, **12**, e0185588.