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

Ian Durbach[1,2,*], Rishika Chopara[3], David L. Borchers[1,2], Rachel Phillip[1], Koustubh Sharma[4], and Ben C. Stevenson[3]

[1]center 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]center for Statistics in Ecology, the Environment and Conservation, Department of Statistical Sciences, University of Cape Town, South Africa

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

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

[*]Corresponding author: indurbach@gmail.com

Appendix A Bayesian models

Results presented in Section 4 in our manuscript 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 conclusions are not simply a consequence of adopting a classical approach. We have reproduced Figures 4 and 5 below.

In Section A1 we describe our Bayesian models, and in Section A2 we summarize our results.

A1 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 predicted AC location surfaces, and a model with inhomogeneous density characterized by a log-linear relationship with a spatial covariate to estimate expected AC density surfaces.

We fitted our models in NIMBLE (de Valpine, Turek, Paciorek, Anderson-Berman, Temple Lang, and Bodik, 2017; Turek, Milleret, Ergon, Brøseth, Dupont, Bischof, and de Valpine, 2021) using data augmentation (Tanner and 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 centers, 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 'exists' in a given MCMC iteration. Rather than directly estimating N, the population size, we estimate the data augmentation parameter, ψ, the 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 $\sum_{i=1}^{M} z_i$. A sample from the posterior for animal density can be obtained by dividing each estimate of N by the area of the survey region. Further details on data augmentation can be found in Kéry and Schaub (2012, pp. 139–157).

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 centers were given a uniform prior distribution over the survey region and the data augmentation parameter was given a uniform prior from 0 to 1. For the inhomogeneous density model,

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 11 000 MCMC iterations, where we set M to be equal to 300. We also used an adaptation interval of 1000, and discarded 1000 iterations as burn-in.

When fitting each inhomogeneous density model, we ran 101 000 MCMC iterations, and used a value of 300 for M. We didn't use an adaptation interval, and discarded 1000 iterations as burn-in.

A2 Results

We created trace plots for all parameters across all models, and none of them indicated a lack of convergence. Although we do not present them here for brevity, the point estimates (calculated using the posterior mean) of all parameters were similar to those obtained via maximum likelihood models fitted to the same data. In addition, the 95% credible intervals for the parameters were similar to the 95% confidence intervals found via maximum likelihood models.

The plots based on our Bayesian models fitted via MCMC (Figures A1 and A2, respectively) were qualitatively similar to those based on maximum likelihood models (respectively, Figure 4 and Figure 5). A potential explanation for any subtle differences is that our Bayesian plots are constructed based on entire posterior distributions, whereas the maximum likelihood alternatives only use point estimates.

Appendix B Uncertainty surfaces

Here we present surfaces that convey the uncertainty present in the expected AC density surface estimates shown in A1 and A2. The left and right columns

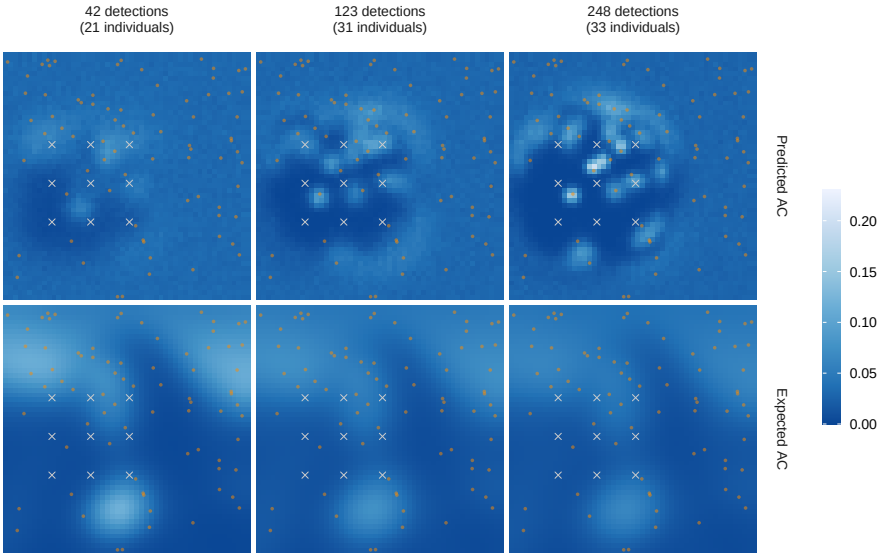

Figure A1: A version of Figure 4 from the manuscript based on our Bayesian models fitted via MCMC.

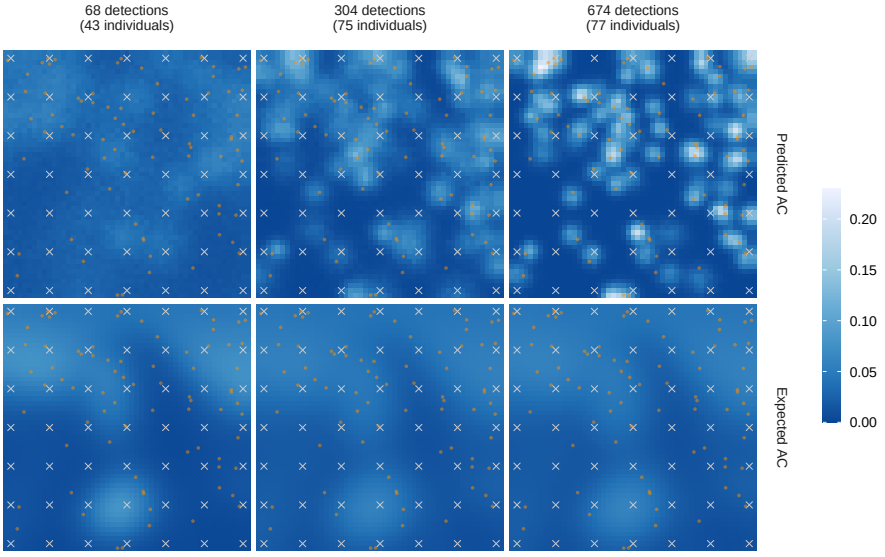

Figure A2: A version of Figure 5 from the manuscript based on our Bayesian models fitted via MCMC.

of Figures B3 and B4 show the lower- and upper-limits of 90% credible intervals, using the 5% and 95% quantiles of the posterior distribution for the expected AC density in each pixel. The middle column shows the expected AC density surface estimates from A1 and A2, using the posterior mean for each pixel. For a frequentist analysis, calculating measures of uncertainty (standard errors and confidence intervals) can be achieved using the delta method or a bootstrap, for example.

We also considered uncertainty in predicted AC location surface estimates. For a Bayesian analysis, we can approximate the posterior distribution of a pixel's AC density by counting the number of ACs within the pixel on each MCMC iteration, and dividing by the area of the pixel. For a frequentist analysis, defining and calculating uncertainty is more complicated: both the delta method and the bootstrap are designed to calculate variance of estimates across multiple realizations of the AC point process, but predicted AC location surfaces are specific a single realization.

Here we consider both the posterior standard deviation the coefficient of variation (CV) as a relative measures of uncertainty. The CV is calculated by dividing the standard deviation of the posterior distribution by the mean. In Figure B5, we display posterior standard deviations and CVs for the realized AC density surface estimate shown in the top-left panel of Figure A1. We calculated these measures of uncertainty for three different pixel sizes: squares with edges of 1 distance unit, 2 units, and 5 units, respectively.

Spatial patterns in posterior standard deviation reflected the estimated surface, because pixels that tend to have higher counts of ACs in pixels inherently have more variation across iterations. The CV measures relative rather than absolute uncertainty, thus correcting for differences in uncertainty across pixels caused by variation in the posterior mean. The spatial patterns in CV were the opposite of those for the posterior standard deviation, with highest levels of relative uncertainty in the pixels with the smallest estimates of AC location density.

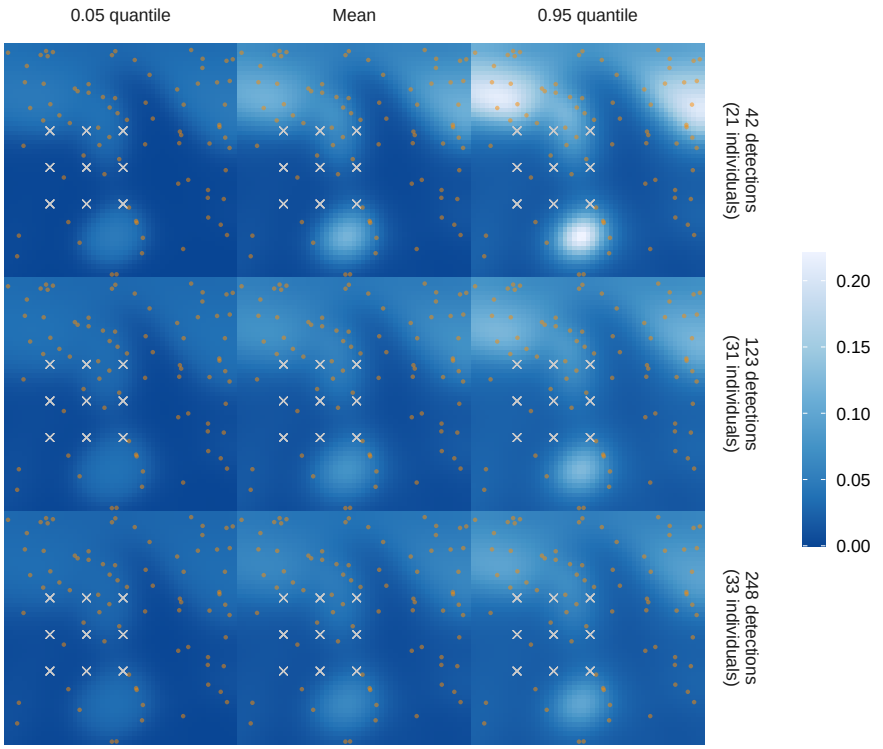

Figure B3: Plots visualising the uncertainty present in the expected activity center surfaces from Figure 4.

We found that uncertainty in predicted AC location density was sensitive to pixel size. Using smaller pixels results in larger uncertainty due to smaller samples of ACs within the pixel across the MCMC iterations, while for larger pixels the relative number of ACs per unit area within a pixel is more stable.

We do not get the same result for uncertainty in expected AC density surface estimates, because these deal with the expected (rather than observed) number of ACs per unit area in the pixel. The CVs for expected AC density estimates are substantially lower than for predicted AC location estimates, because estimating the mean of a distribution can be achieved more accurately than predicting a realization from that distribution. Predicting outcomes of random variables requires that we account for the variance of the distribution itself, in addition to variation due to uncertainty in estimating the mean.

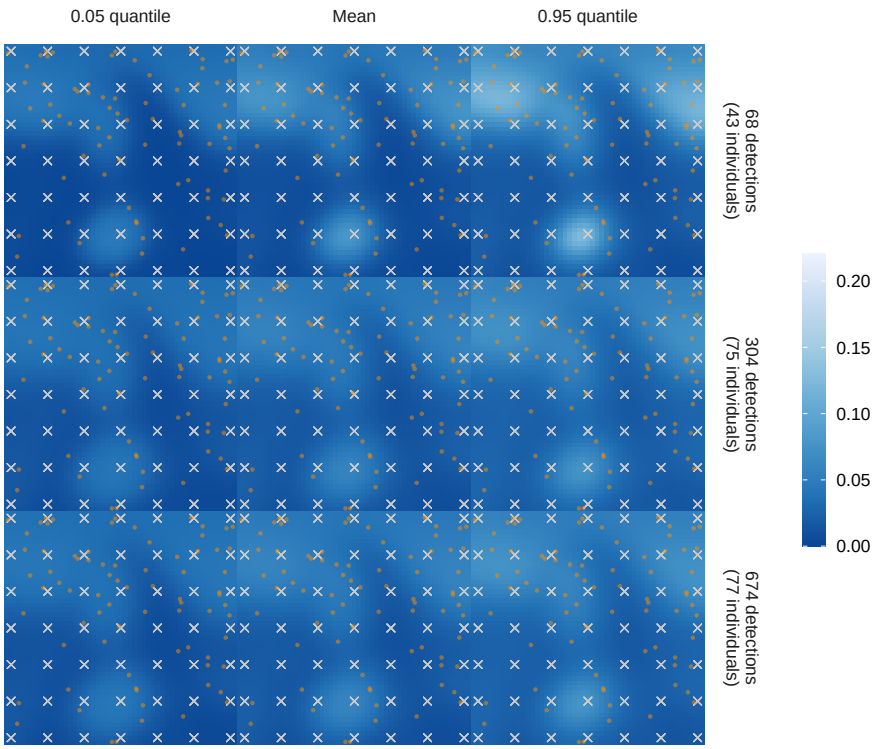

Figure B4: Plots visualising the uncertainty present in the expected activity center surfaces from Figure 5.

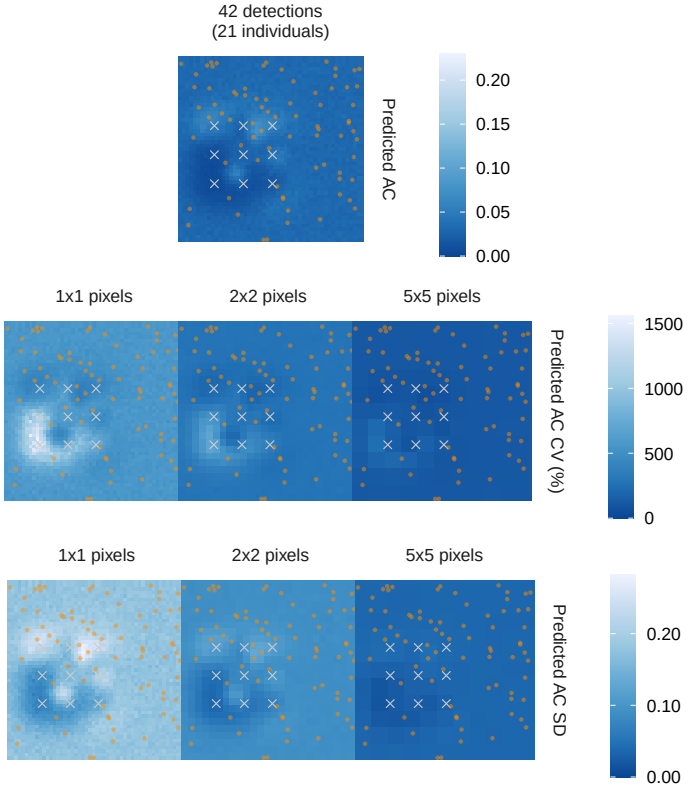

Figure B5: Plots visualising how the CV and posterior standard deviation values change as pixel size changes, using the the leftmost predicted AC location surface in Figure 4.

References

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

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

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

Turek, D., Milleret, C., Ergon, T., Brøseth, H., Dupont, P., Bischof, R., and de Valpine, P. (2021). Efficient estimation of large-scale spatial capturerecapture models. *Ecosphere* **12,** e03385.