

## **Project 3: Dynamic Occupancy Model - Sutherland Water Voles**

### **Introduction**

The study of metapopulation dynamics is crucial to understanding and mitigating threats to biodiversity resulting from habitat fragmentation (*Hanski, I., 1998*). This concept refers to the state of a species' occupied habitats being broken up into multiple patches due to the existence of a landscape matrix unsuitable for reproduction, such that each subpopulation may be modelled itself as a population related to the others by the processes of dispersal and demographic stochasticity. The two fundamental parameters driving the dynamics of such systems are the probability of patch extinction, whereby none of the individuals in a subpopulation survive, and the probability of colonisation, whereby individuals disperse from one patch to another in search for better conditions. Thus, patch occupancy dynamics are predicted to be governed by an 'area-isolation paradigm', whereby extinction probabilities are inversely related to patch size and colonisation probabilities are positively related with the connectivity of that patch to others. In this investigation, we developed a dynamic spatial occupancy model describing a metapopulation of water voles (*Arvicola amphibius*), a riparian species exhibiting natural population fragmentation in Sutherland, Scotland. Here we present the methods and findings of our investigation undertaken to test the predictions of spatially realistic metapopulation theory.

### **Methods**

The data described a network of a hundred and fourteen patches which were each surveyed between one and four times during the breeding season over a period of four years from 2009 to 2012. The detection-non-detection data is based on discovery of at least one 'latrine' (faecal deposit) at each patch during a survey. We were provided with the number of surveys for each patch-year combination as well as how many of these resulted in water vole detection. We used Markov-Chain Monte Carlo (MCMC) simulation (*Plummer et al., 2006*) to draw samples from the posterior distribution of the developed probabilistic model in order to obtain estimates for the parameters of interest.

Our state-space model consisted of three parts - the priors and constraints, likelihood of the state process, and likelihood of the observation process. The state process was constructed with the following likelihoods:

$$\begin{aligned}
 z_{i,1} &\sim \text{Bernoulli}(\psi_1) \\
 z_{i,t} | z_{i,t-1} &\sim \text{Bernoulli}(\psi_{i,t}) \\
 \psi_{i,t} &= (1 - z_{i,t-1})\gamma_{i,t-1} + z_{i,t-1}(1 - \varepsilon_{i,t-1}) \\
 \text{logit}(\gamma_{i,t}) &= \beta_0^\gamma + \beta_1^\gamma \times \text{Connectivity}_i \\
 \text{logit}(\varepsilon_{i,t}) &= \beta_0^\varepsilon + \beta_1^\varepsilon \times \text{Length}_i.
 \end{aligned}$$

$z_{i,1}$  represented the occupancy state (which either equals 1 when occupied or 0 when not) for each patch in the first year, and was modelled as a single binary trial with an initial occupancy probability  $\psi_1$  uniform across all patches. The evolution of the occupancy state was modelled as Markovian and distributed as a binary trial, but with each patch-year combination having a different success probability  $\psi_{i,t}$ . These probabilities were dependent on the colonisation and extinction probabilities from the previous year, which were also allowed to vary across both patch and year. To test the 'isolation' prediction, we modelled the colonisation probabilities as linearly predicted by the explanatory variable  $Connectivity_i$  on the logit link scale. Connectivity

score is given by  $Connectivity_i = \sum_{i \neq j} \exp(-0.33 \times d_{ij})$ , where  $d_{ij}$  represents the Euclidean distance in km between patches  $i$  and  $j$  (Sutherland C., et al., 2014). Similarly, to test the 'area' prediction, we modelled the extinction probabilities as linearly predicted by the explanatory variable  $Length_i$ , a measure of patch size, again on the logit scale. Followingly, there were two parameters to be determined in each equation. The observation process was constructed using the following likelihoods:

$$\begin{aligned} q_{i,t} &\sim Normal(\mu_p, \sigma_p^2) \\ \text{logit}(p_{i,t}) &= q_{i,t} \\ y_{i,t} | z_{i,t} &\sim Binomial(J_{i,t}, p_{i,t} \times z_{i,t}). \end{aligned}$$

The number of surveys for which there was positive water vole detection was modelled as a Binomial series of  $J_{i,t}$  trials, where  $J_{i,t}$  is the number of trials for a given patch in a given year, each with probability given by the product of the specific patch-year detection probability and the latent patch-year occupancy state (either 0 or 1). A logit link was again used to model the detection probabilities, and dependent on a Normal distribution whose parameters were to be estimated in the analysis.

Diffuse priors were used for all parameters used in the likelihood modelling as we had no *a priori* information on which to base hypotheses. Normal distributions with mean 0 and standard deviation 0.01 were used as the priors for the coefficients  $\beta_0^y, \beta_1^y, \beta_0^e$ , and  $\beta_1^e$ . The mean of the patch by year detection probabilities,  $\mu_p$ , was also modelled with a Normal prior, whereas its standard deviation was given by a Uniformly distributed prior. The analysis was conducted entirely in R software (R Core Team, 2021) using the Just Another Gibbs Sampler (JAGS) package (Depaoli et al., 2016) for Bayesian analysis Using Gibbs Sampling (BUGS) language (Lunn et al., 2000).

For the MCMC simulation, three Markov chains each of ten thousand iterations were used to run the analysis with four thinning samples. In order to achieve better convergence for our simulation, burn-in of some iterations was necessary. We chose a burn-in value of a thousand

iterations based on inspection of the trace plots with zero burn-in and deemed this to be sufficient to maximise probability of the chains' convergence.

Once we had obtained the state-space model output, we assessed the fit using two metrics for posterior predictive checking - chi-squared statistics, given by  $\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ , and Freeman-Tukey fit statistics, given by  $FT = \sum (\sqrt{\text{observed}} - \sqrt{\text{expected}})^2$ . Data simulated based on the model fit were compared to the observed data and discrepancies between the two were calculated by these metrics to determine the goodness of our model fit.

## Results

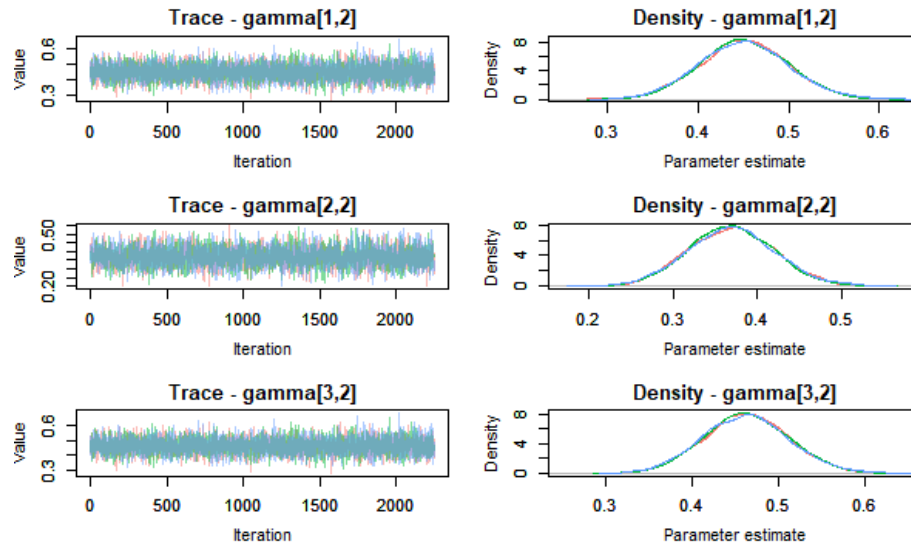
Our model output gave estimates, standard deviations, and 95% confidence intervals for the  $114 \times 3 = 342$   $\gamma_{i,t}$  and  $\varepsilon_{i,t}$  coefficients, and single coefficients for the other parameters. The derived parameters of proportional occupancy for the patches in each year are also estimated. Estimates for the parameters other than the are shown in the table below.

| Parameter  | Mean   | Sd    | Lower  | Medium | Upper  | Rhat | n.eff |
|------------|--------|-------|--------|--------|--------|------|-------|
| beta0g     | -1.033 | 0.345 | -1.734 | -1.021 | -0.384 | 1.00 | 6750  |
| beta1g     | 0.828  | 0.302 | 0.296  | 0.804  | 1.493  | 1.00 | 6750  |
| beta0e     | -1.681 | 0.357 | -2.411 | -1.667 | -1.021 | 1.00 | 2586  |
| beta1e     | 0.464  | 0.346 | -0.251 | 0.469  | 1.138  | 1.00 | 6750  |
| mu_p       | 1.026  | 0.115 | 0.806  | 1.024  | 1.255  | 1.00 | 4564  |
| sig_p      | 0.535  | 0.239 | 0.141  | 0.516  | 0.963  | 1.03 | 139   |
| propocc[1] | 0.516  | 0.019 | 0.482  | 0.518  | 0.561  | 1.00 | 725   |
| propocc[2] | 0.599  | 0.021 | 0.561  | 0.596  | 0.649  | 1.01 | 278   |
| propocc[3] | 0.656  | 0.018 | 0.623  | 0.658  | 0.693  | 1.00 | 747   |
| propocc[4] | 0.698  | 0.021 | 0.667  | 0.693  | 0.746  | 1.00 | 711   |

**Figure I:** summary of the parameter estimates based on the posterior distributions

The  $\hat{R}$  values for the parameters shown in the posterior summary above, as well as for each colonisation probability and extinction probability in their respective estimated matrices, are all crucially less than 1.1, which indicates that the between- and within-chain estimates for the parameters converged. This results from good mixing of the MCMC chains. From inspecting the trace and density plots output by the model for each estimated parameter, it appears that the MCMC explored the parameter space efficiently. This observation is based on the 'grassy' appearance of the trace plots, meaning no chain's signal can be discerned from the others', as well as the three smooth overlapping density curves for each parameter, representing the posterior distributions estimated by each chain. This provides visual reassurance of the model's convergence and allows reasonable confidence in the parameter results. Example trace and

density plots for the estimates of the colonisation probabilities in patches 1, 2 and 3 in the second year of surveying are shown in Figure II.

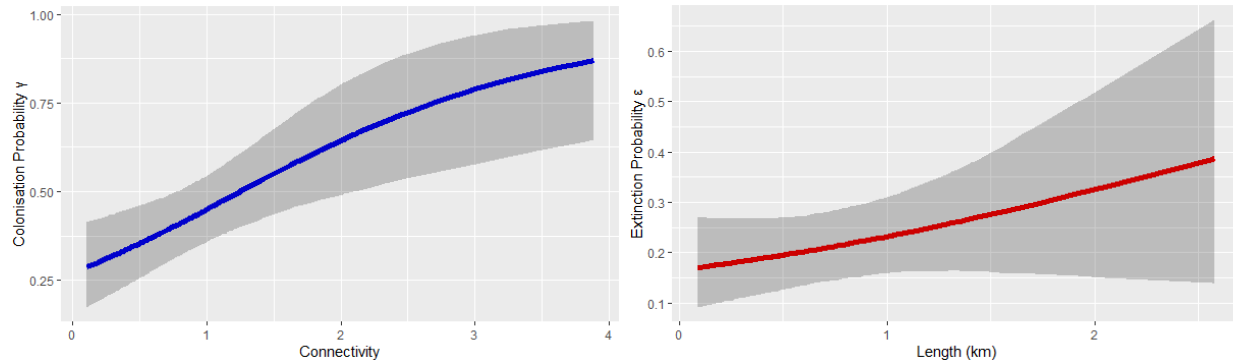


**Figure II:** trace and density plots for the posterior distributions of the parameters  $\gamma_{1,2}$ ,  $\gamma_{2,2}$ , and  $\gamma_{3,2}$  as output by the model. The trace plots show grassy appearance and the density plots show smooth Normally distributed curves which overlap, demonstrating model convergence.

Based on Figure I, we see that the proportion of the hundred and fourteen patches which are occupied (*i.e.* patches for which the occupancy state  $z_{i,t} = 1$ ) increases for each successive survey year, from 0.516 in the first survey year (2009) to 0.698 in the last survey year (2012). We obtain equations for the patch by year colonisation and extinction probabilities respectively as:

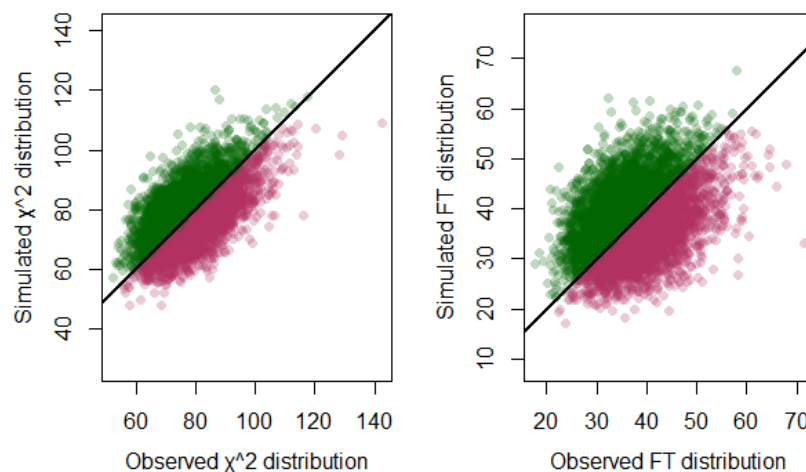
$$\begin{aligned} \text{logit}(\gamma_{it}) &= -1.033 + 0.828 \times \text{Connectivity}_i \\ \text{logit}(\varepsilon_{it}) &= -1.681 + 0.464 \times \text{Length}_i. \end{aligned}$$

The positivity of  $\beta_1^\gamma$  implies that as the connectivity of a patch increases, so too does the odds of water vole colonisation - for a unit increase in the connectivity score, the odds of colonisation increases by a factor of  $\exp(0.828) = 2.29$ . This is in accord with our expectation based on the isolation prediction of metapopulation theory. Comparably, the positive  $\beta_1^\varepsilon$  coefficient implies that as the length of a given patch increases, so too does the odds of extinction - for an increase of 1 km in patch length, the odds of extinction increases by a factor of  $\exp(0.464) = 1.59$ . This result, however, is not in line with our expectation based on the area prediction, which has extinction probability decreasing with increasing patch length. The two relationships are visualised in Figure III.



**Figure III:** plots of the colonisation and extinction probabilities as related to their respective explanatory variables, patch connectivity and patch length. The means of each parameter are represented by the coloured lines and the 95% confidence bounds are displayed as the surrounding grey regions.

Finally we assessed the model's fit to be able to gauge the accuracy of our results. We plotted both the chi-squared and Freeman-Tukey statistics for the observed vs the simulated data to compare how closely the simulated data fit with the observed data. Based on Figure IV, the model does seem to provide a good fit for the data judging from the fact that the observed fit statistics correspond closely with those from the simulated data. The Bayesian p-values, *i.e.* the probabilities that the observed fit statistic was more extreme than that from the simulated data, were calculated as  $Bpv = 0.456$  based on the chi-squared discrepancy metric, and  $Bpv = 0.482$  based on the Freeman-Tukey metric. This gives numerical confirmation of the goodness of fit of our model, as both p-values are not extreme, lying close to 0.5.



**Figure IV:** scatter plots of the two fit statistics -  $\chi^2$  and Freeman-Tukey - obtained from the observed and simulated data. The black lines represent equality of the statistics based on the observed and simulated data; the points are maroon when the observed statistic is greater than the simulated statistic, and green when vice versa.

## Discussion

Above we have described the motivation for the investigation, the process of constructing the dynamic occupancy state-space model, evaluation of the model's fit with regard to the observed data, and the model's estimation for the parameters we wished to monitor. Based on our assessment of the model's MCMC convergence through the trace plots and  $\hat{R}$  measures, we were able to have confidence in the robustness of our output. The posterior predictive check for evaluation of the model fit was also assuring of its accuracy in relation to the data observed, as it showed the simulated data closely matching the observations. Based on our estimate for the matrix of patch occupancy probabilities, it was found that the proportional occupancy of patches in the water vole metapopulation grew with each following year, which presents no concern as relates to the prevention of biodiversity loss.

In the context of the primary objective of the analysis, *i.e.* to test the predictions of spatially realistic metapopulation theory, we obtained results in varying agreement with our expectation. The isolation prediction was confirmed by our analysis, which showed the colonisation probability as being positively linked to patch connectivity. In relation to the area prediction, however, it was found that the length of a patch was not inversely related to its probability of extinction, but rather positively related. This suggests that as the size of a patch increases, the water vole subpopulation specific to that patch faces greater risk of extinction. A possible explanation may have to do with population size increasing non-linearly with patch length such that there is increasing competition for resources in increasingly large patches, reducing each individual's probability of survival and therefore also increasing the probability of extinction within the patch. There may also be confounding factors related to patch length, such as increased predation, which affect this process. However, further study of the metapopulation area prediction is needed to shed more light on this issue.

## References

- Hanski, I., (1998) ,  
“Metapopulation Dynamics”,  
*Nature*, 396(6706), pp. 41-49,  
Available at: <https://doi.org/10.1038/23876>
- Sutherland, C. S., Elston, D. A., and Lambin, X. (2014),  
“A demographic, spatially explicit patch occupancy model of Metapopulation Dynamics and Persistence”,  
*Ecology*, 95(11), pp. 3149-3160,  
Available at: <https://doi.org/10.1890/14-0384.1>
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006),  
CODA: Convergence Diagnosis and Output Analysis for MCMC,  
*R News*, vol 6, 7-11  
Available at: [https://cran.r-project.org/doc/Rnews/Rnews\\_2006-1.pdf#page=7](https://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf#page=7)
- R Core Team (2021),  
R: A Language and Environment for Statistical Computing,  
Vienna, Austria.  
Available at: <https://www.R-project.org/>
- Depaoli, S., Clifton, J. P., & Cobb, P. R. (2016),  
Just Another Gibbs Sampler (JAGS): Flexible software for MCMC implementation,  
*Journal of Educational and Behavioral Statistics*, 41(6), 628–649.
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000),  
WinBUGS - a Bayesian Modelling Framework: concepts, structure, and extensibility.  
*Statistics and Computing*, 10:325–337.

## Code Supplement

Please see separately attached R file “MWPD Assignment III.R” for fully commented code describing the model.