Supplementary Material 3:

Bayesian occupancy modelling of benthic Crustacea and the recovery of the European spiny lobster, *Palinurus elephas*

Jackson, A.C.

Traceplots for Monte Carlo Markov Chains

For *Palinurus elephas* from Cornwall (which showed interesting trends in occupancy, see Fig. 4) plots of each MCMC chain for occupancy and detectability for every fifth year in the series (Fig S3.1-4) were inspected for patterns that might indicate issues with the data or model. Plots of the first 300 iterations (Fig. S3.1-2) were used to see detail in the progression of the algorithm, whereas the plots of 5000 iterations (Fig S3.3-4) give an impression of the shape of the plots for the entire run. Each parameter showed regular state-changes, with no evidence of bimodality in density plots. In each case, the three traces (red, black or blue) showed low serial correlation and no persistence of burn-in, indicating that the *a priori* distribution was well calibrated and that all parameters had converged.

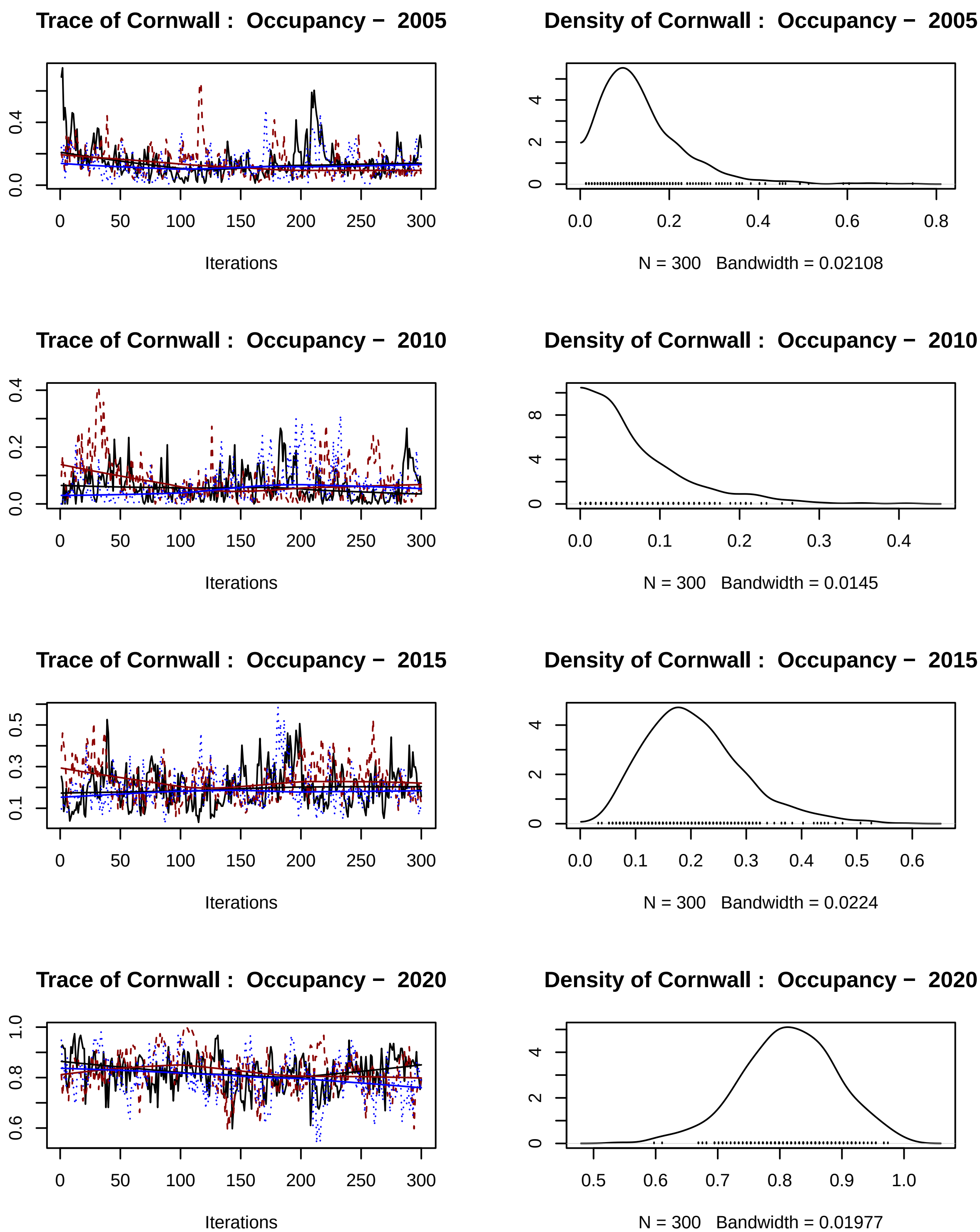


Figure S3.1. Traceplots of three MCMC chains (red, black, blue) for estimates of occupancy by *Palinurus elephas*. Only 300 iterations are shown such that fine detail is visible.

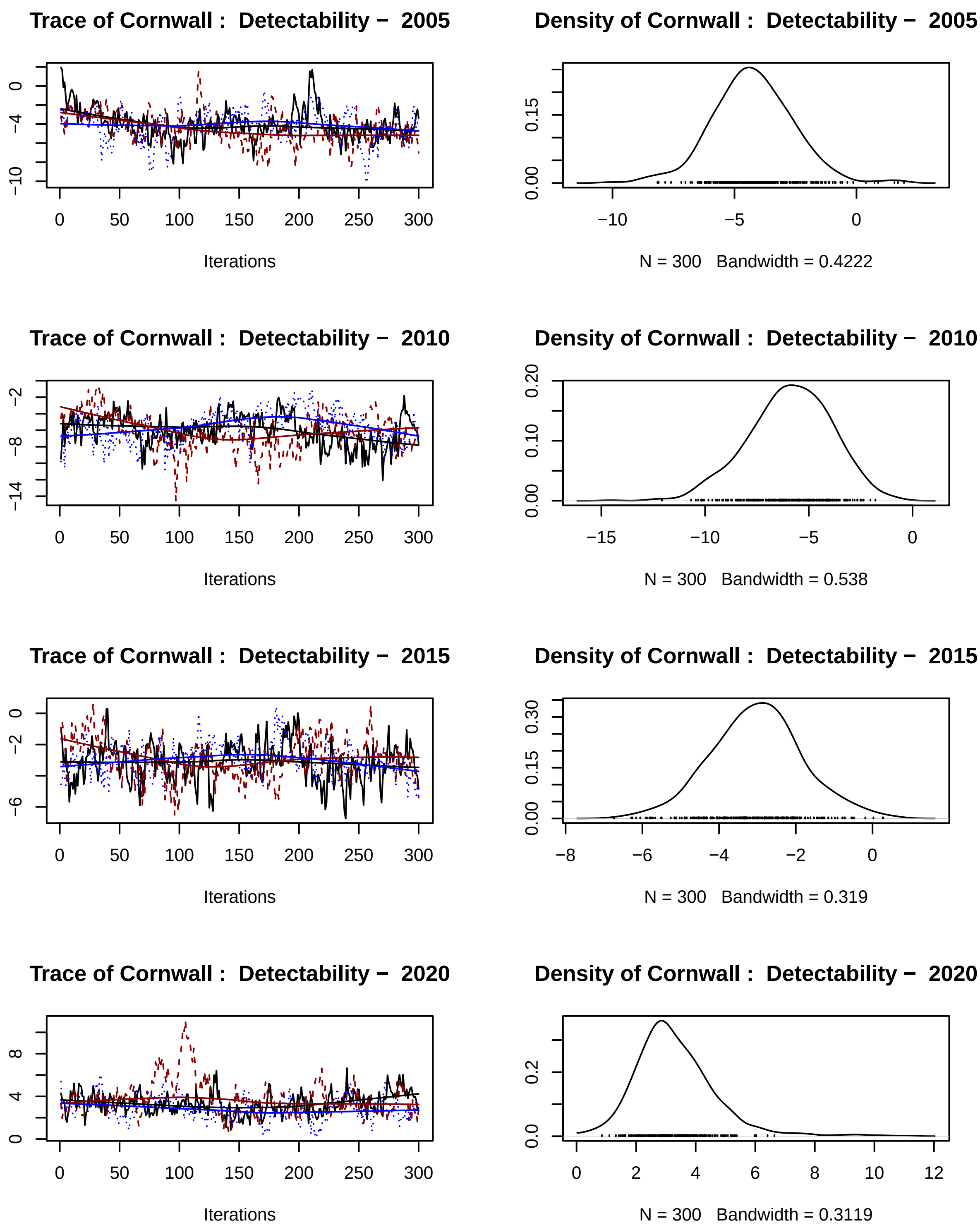


Figure S3.2. Traceplots of three MCMC chains (red, black, blue) for estimates of detectability for *Palinurus elephas*. Only 300 iterations are shown such that fine detail is visible.

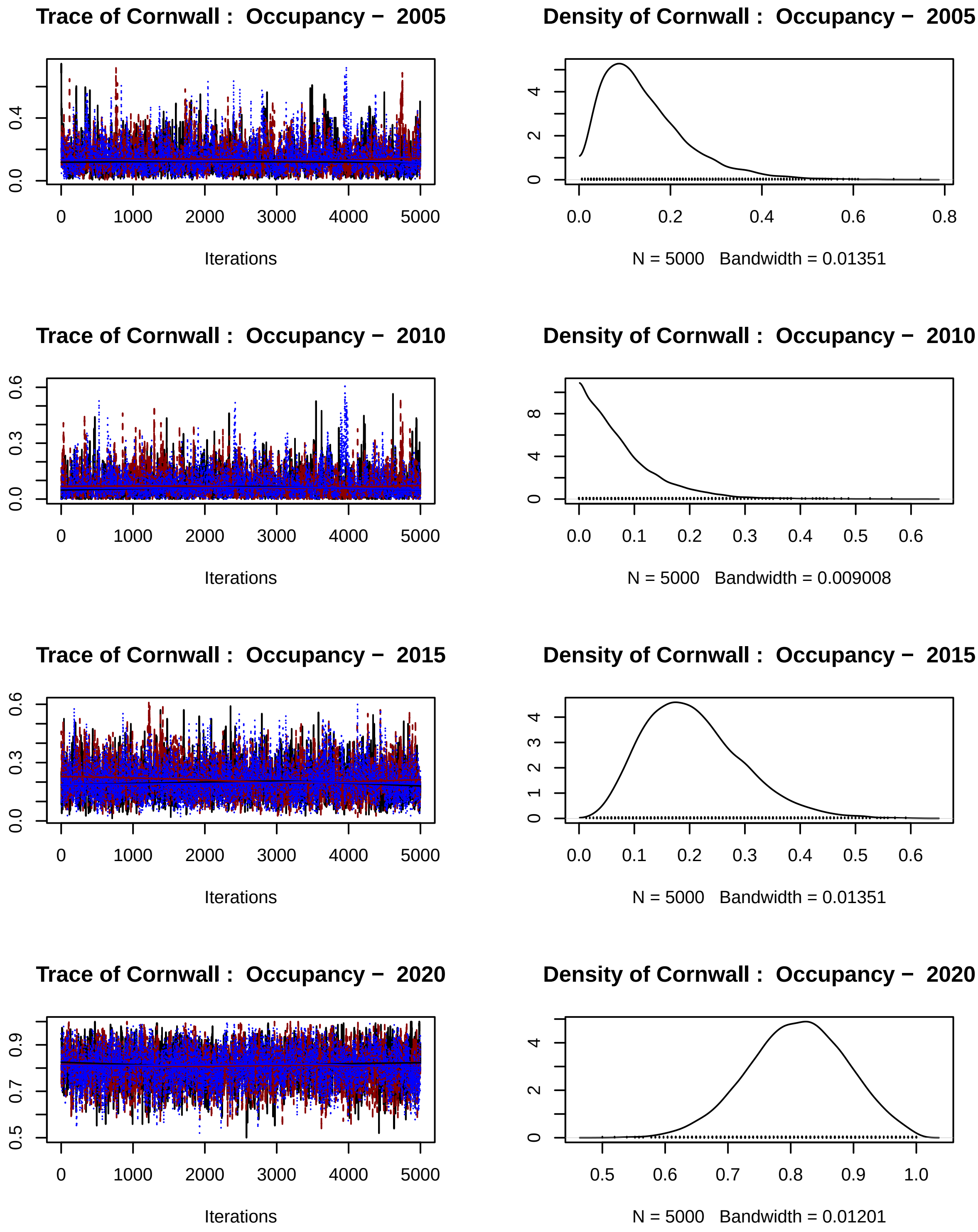


Figure S3.3. Traceplots of three MCMC chains (red, black, blue) for estimates of occupancy by *Palinurus elephas*. 5000 iterations are shown to illustrate coarse patterns in the algorithm.

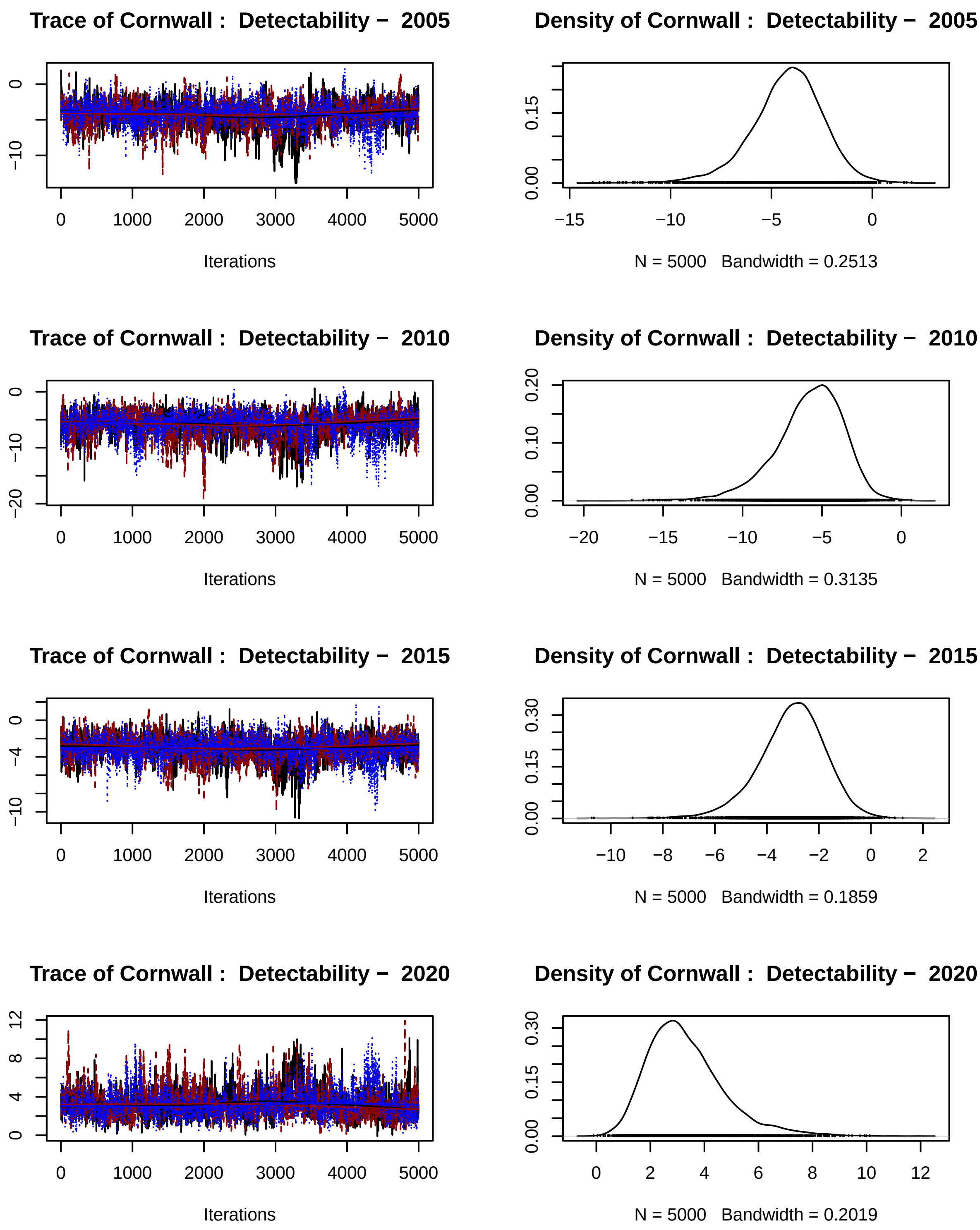


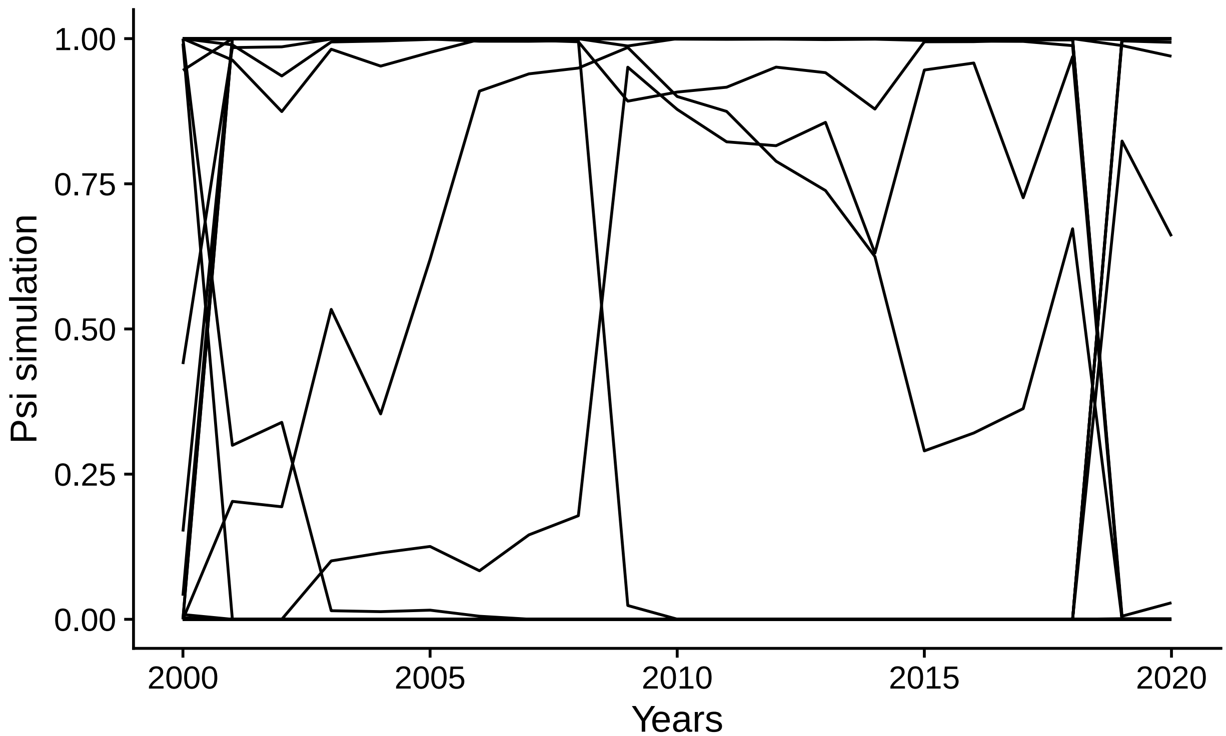
Figure S3.4. Traceplots of three MCMC chains (red, black, blue) for estimates of detectability for *Palinurus elephas*. 5000 iterations are shown to illustrate coarse patterns in the algorithm.

Graphical Posterior Predictive Checks

Graphical checks of the posterior predictions are often recommended in order to assess the suitability of the model selected. One way to do this for the model used here would be to use posterior samples from the parameters of the random walk ($b_1,\ \mu_b,\ \sigma_b^2$) to generate values of the occupancy probabilities (ψ_{it}). Within the sparta package, the parameters of the random walk ($b_1,\ \mu_b,\ \sigma_b^2$) are not automatically monitored or stored as posterior samples. Thus, it is not straightforward to generate simulations of occupancy from the posteriors of the year effects. Although it may be possible to generate the required posteriors using the additional.parameters argument of the model, no information is provided in the sparta github (https://github.com/BiologicalRecordsCentre/sparta) about how to do this. Early demonstrations of the sparta tool did, however, provide some convincing evidence about the broad applicability of the model recommended in Outhwaite et al., (2018, 2019).

The prior for the year effects with a random walk (as used here) is widely dispersed (e.g. Equations 9 & 10: b_1 with variance = 10^4) with the intention of it being uninformative (leading to a uniform distribution of occupancy probabilities). As demonstrated by Outhwaite *et al.*, (2018; Appendix D), the variance of the prior of the year effects under the random walk model grows without limit as t increases and hence the variance and distribution of b_t depends on t. When b_t is transformed to ψ_{it}, it leads to a clear, u-shaped distribution of probabilities of occupancy (Figure S3.5), i.e. that at most points during the time-series, occupancy estimates from the year effects alone are almost 1.00 or almost 0.00 (i.e. they are not uniform and consequently, not uninformative). Thus, under this model, it is not possible to specify a prior that gives a uniform distribution of occupancy probabilities. This raises concerns that the prior may exercise excessive control over the results and about the suitability of the model.

To help allay these concerns, Outhwaite *et al.*, (2018, Appendix D) went on to demonstrate that the prior chosen for the random walk is able to adapt quickly when provided with small amounts of data. Their further exploration of the relative merits of using the base model or the random walk with either uniform or half-Cauchy priors for the ψ_{i1} revealed that with the uniform priors, estimates were drawn towards a 0.5 probability of occupancy when there were few or no data early in the time-series and could give confident, but unrealistic estimates of occurrence at the start. In contrast, where priors for year effects were specified by a random walk with half-Cauchy distribution of variance, estimates of occupancy where there were few data, were more credible (Outhwaite *et al.*, 2018; Appendix D). Under all scenarios encountered by those authors, the random walk model with the u-shaped prior did in fact adjust rapidly to small amounts of information and produced intuitive and reasonable outputs. This is of course, not to say that more appropriate models do not exist for different scenarios, just that they have yet to be formulated within sparta and evaluated on citizen-science data.



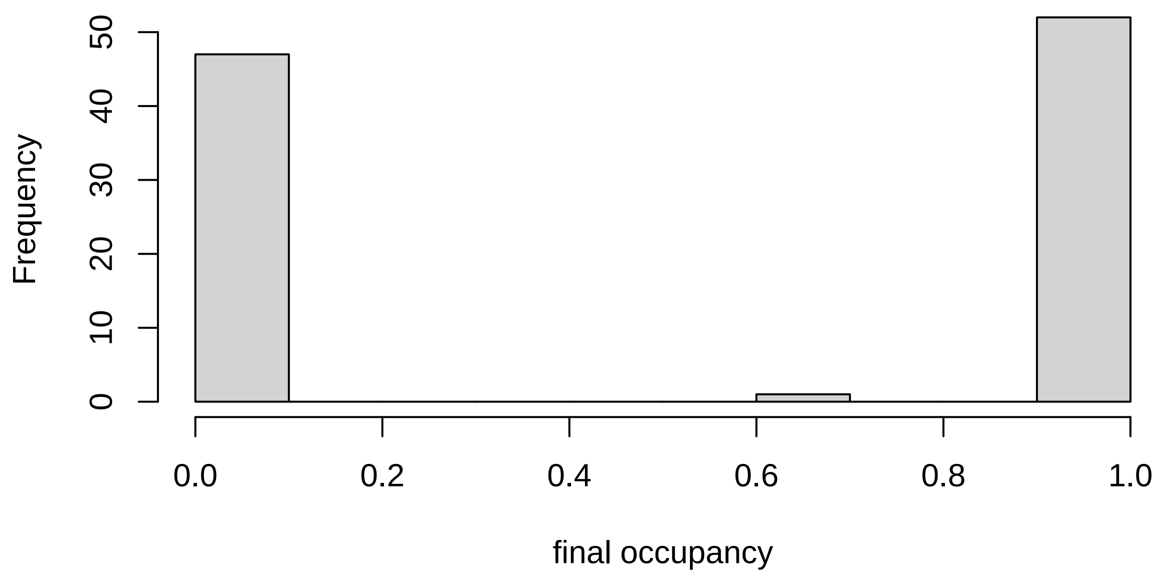


Figure S3.5 Values of psi simulated from priors (n = 100 iterations) and histogram of probabilities in the final time-step (year 21). Many of the iterations are not visible in the upper plot as they fall along the values of 1.00 or 0.00.