

Ecography

**E7190**

Catterall, S., Cook, A. R., Marion, G., Butler, A. and Hulme, P. E. 2011. Accounting for uncertainty in colonisation times: a novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. – *Ecography* 34: xxx–xxx.

**Supplementary material**

## Appendix 1

### Details of the MCMC inference algorithm

Here we provide some more details of the MCMC algorithm. The first step is to define prior distributions for each model parameter. The prior for  $\lambda$  is denoted  $P(\lambda)$ , with similar notation for other model parameters. When we applied the inference algorithm to the distribution data for giant hogweed, we took  $P(\lambda)$  to be uniform on the interval  $(0, 10)$ ,  $P(\alpha)$  to be uniform on  $(-40, 40)$ ,  $P(\tau)$  to be uniform on  $(-40, 40)$  and  $P(\beta_k)$  to be uniform on  $(0, 10)$  for  $k = 2, 3, \dots, 10$ . The suitability parameter for sea ( $\beta_1$ ) was fixed at zero.

Now we describe the procedure for simulating the Markov chain whose stationary distribution is the posterior distribution we are interested in. An iteration of the algorithm comprises sequential proposed updating of all model parameters and unknown colonisation times. A superscript  $(m)$  is used to indicate the value of a quantity at the  $m$ 'th iteration of the MCMC algorithm. The algorithm is structured as follows.

(1) Read in snapshot distribution data. Initialise model parameters with arbitrary values. Initialise each unknown colonisation time in  $H$  (the set of unobserved colonisation times in the time interval  $[t_1, t_N]$ ) uniformly at random on the interval of possible times  $[t_{n-1}, t_n]$  discussed in Methods. Compute initial likelihood  $P(H|\theta, C(t_1))$ .

(2) Repeat the following sequence of steps a large number of times – we usually used at least 25000 iterations.

Suppose we are at iteration  $m$ .

(a) Evaluate the current likelihood  $L$ . Propose a new value  $\lambda^*$  for the parameter  $\lambda$  using the proposal distribution  $q(\lambda^{(m)}, \lambda^*)$ . Recalculate the likelihood using the proposed  $\lambda^*$ , giving value  $L^*$ . With

probability  $\min\left(1, \frac{L^*P(\lambda^*)q(\lambda^*, \lambda^{(m)})}{LP(\lambda^{(m)})q(\lambda^{(m)}, \lambda^*)}\right)$  accept the proposed change, so that  $\lambda^{(m+1)} = \lambda^*$  and the

likelihood is updated to  $L^*$ . Otherwise, we take  $\lambda^{(m+1)} = \lambda^{(m)}$  and the likelihood is unchanged.

(b) Repeat step (a) for the other model parameters  $\alpha_1, \alpha_2, \dots, \alpha_K$ .

(c) Repeat step (a) for each of the unknown colonisation times  $t$  in  $H$ , the only difference being that the

acceptance probability is  $\min\left(1, \frac{L^*q(t^*, t^{(m)})}{Lq(t^{(m)}, t^*)}\right)$ .

The initial iterations of step (2) are normally discarded as 'burn-in' so as to allow for convergence of the Markov chain, removing any effect of the arbitrary initial parameter choices. A decision must be made regarding the actual length of the burn-in period. Figure A.1 shows trace plots and marginal posterior histograms for the dispersal parameter  $\lambda$  and the broadleaf suitability parameter  $\beta_4$  when the parameter estimation algorithm is applied to the 1970 and 2000 giant hogweed distribution maps. For all model parameters, visual inspection of the trace plots suggested convergence after a small number of MCMC iterations. A burn-in of 5000 iterations was felt to be sufficiently long, with 25000 MCMC iterations used in total. Use of the Gelman-Rubin convergence test (Gelman & Rubin 1992) with 16 different sets of starting parameters provided further evidence of convergence.

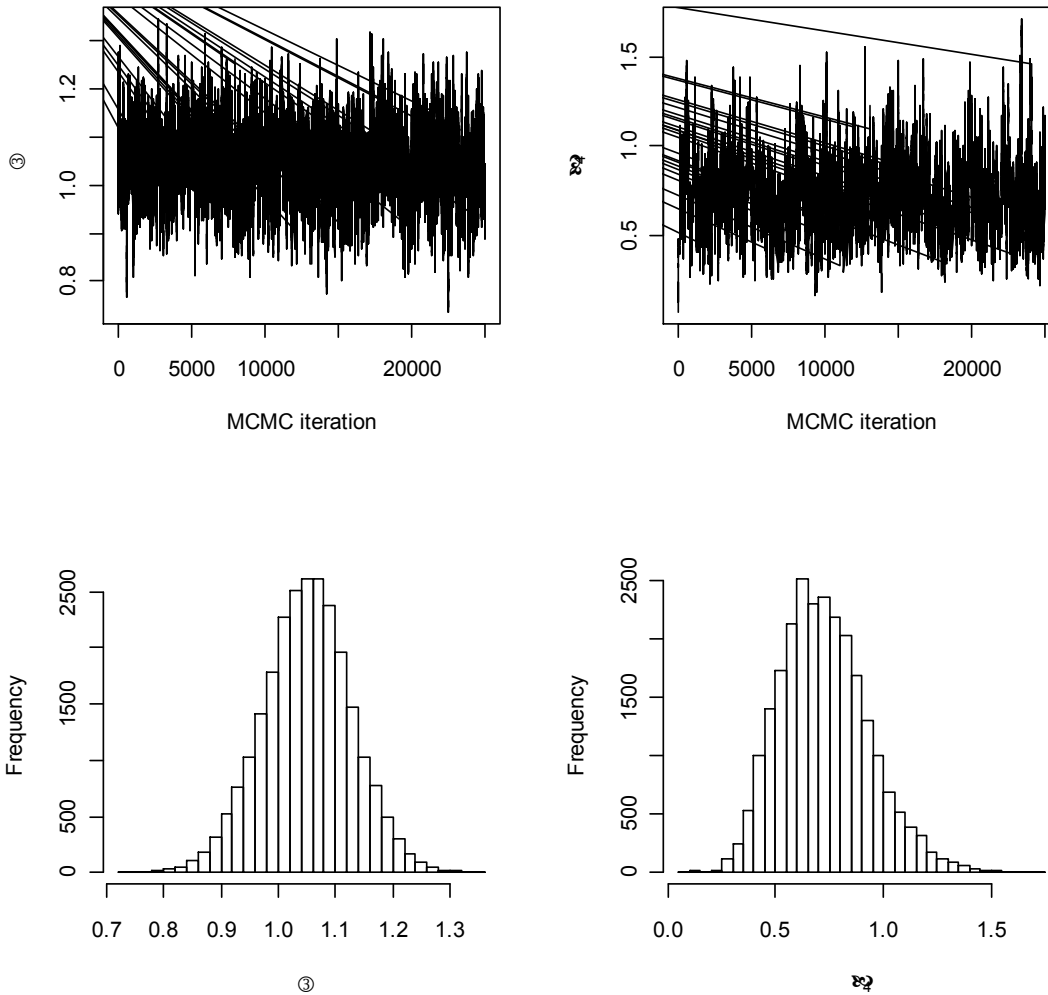


Figure A.1. Top: trace plots for  $\lambda$  (dispersal) and  $\beta_4$  (broadleaf forest suitability) obtained from applying the parameter estimation algorithm to the 1970 and 2000 giant hogweed distribution maps. Bottom: histograms of the marginal posterior distributions for  $\lambda$  (dispersal) and  $\beta_4$  (broadleaf forest suitability) obtained from applying the parameter estimation algorithm to the 1970 and 2000 giant hogweed distribution maps.