

Bayesian Inference for a Complex Ecological System: Viability and Reproduction under Finite Resource Constraints with Incomplete Observations.

C. Jessica E. Metcalf

Nicholas School for the Environment, Duke University, Durham, North Carolina, USA

David A. Stephens

Department of Mathematics and Statistics, McGill University, Montreal, Canada

Mark Rees

Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

Svata M. Louda

School of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA

Kathleen H. Keeler

School of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA

Summary. We address the problem of MCMC analysis of a complex ecological system using a Bayesian inferential approach. We describe a complete likelihood framework for the life history of the wavyleaf thistle, including missing stages and density dependence. We indicate how, in order to make inference on life history transitions involving both missing information and density dependence, the stochastic models underlying each component can be combined to obtain expressions that can be directly sampled. Our approach allows otherwise unobtainable inference on interactions between demographic rates. This innovation and the principles described could be extended to other species featuring such missing stage information, with potential for improving inference relating to a range of ecological or evolutionary questions.

Keywords: Density Dependence; Ecological System; MCMC; Missing Data

1. Introduction

The study of the ecological and evolutionary dynamics of plant populations relies on demographic models based on repeated observations on individuals from natural populations (e.g. Rees *et al.* (2001), Rose *et al.* (2005)). However, key demographic transitions often cannot be directly observed without perturbing the study population, particularly in plant species that reproduce vegetatively, that is, where production of offspring occurs by branching from the taproot, which may take place deep below-ground. Nevertheless, understanding such demographic transitions is critical in a number of fields. Successfully managing invasive plant species requires accurate demographic models, and vegetative reproduction is often a key element to rapid spread of such species (Radosevich *et al.*, 1997). In evolutionary biology, understanding persistence and maintenance of mixed reproductive strategies (i.e. those with vegetative and sexual components) is a fundamental question (Gardner & Mangel, 1999). However, difficulties associated with estimating allocation of finite resources between the different reproductive alternatives, particularly in the presence of unobserved states, mean that evolutionary theory has rarely been aligned with complex real systems.

The fundamental obstacle to statistical inference addressed here is the presence of the missing data relating to the vegetative reproduction. In this paper, we develop a Bayesian

E-mail: cjm29@duke.edu

Address for correspondence: A201 LSRC, Box 90308, Duke University, Durham, NC 27708, USA

Table 1. Longitudinal Record for Typical Ramet: First observed in Year 2 of study, producing 2,1,0 and 1 ramets in Years 2 to 5, and flowering in Year 5 with 2 flower heads.

Year	2	3	4	5
No. Ramets	2	1	0	1
Flower (0=No,1=Yes)	0	0	0	1
No. Flowering Heads	0	0	0	2
Death (0=No,1=Yes)	0	0	0	0

framework to capture all aspects of plant demography, including such missing information, and apply it to an analysis of a large data set relating to the wavyleaf thistle, *Cirsium undulatum*. The model formulation also directly reflects a demographic model which can be used for population projection. Although there is a large body of work relating to the estimation of demographic parameters from the study of marked individuals, including formal statistical (in fact, Bayesian) approaches (Brooks *et al.*, in prep.), explicit linkage to a full demographic model is rare, particularly with connection to unobserved values. Our model and analysis provide such linkage. Inference from our model is performed using a Bayesian framework via Markov chain Monte Carlo.

The paper is organized as follows: the remainder of Section 1 introduces the data considered here, and outlines use of Bayesian approaches in ecological and evolutionary demography to date; Section 2 outlines the probability models and likelihood framework, and presents results from simulated data; Section 3 presents results from analysis of the real data. Conclusions are given in Section 4. Details of notation are given in an Appendix.

1.1. The wavyleaf thistle species and data

Wavyleaf thistle reproduces both vegetatively and sexually, and the resulting offspring can be distinguished in the field. A single genet may consist of several above ground rosettes, referred to as “ramets”, which branch off an underground taproot. Because of the extreme difficulty of identifying genets, in this work we take a single ramet as the focal individual unit. Each ramet persists for a variable number of years until it either dies or flowers. Flowering is generally fatal for a ramet of this species. Each ramet may itself produce vegetative ramets throughout its lifetime. Thus a single longitudinal record for a ramet may typically take the form presented in Table 1. It is of considerable biological interest to understand the nature of each of these reproductive and temporal survival processes, and to understand whether external environmental factors exert any influence because of their ecological and evolutionary implications.

The number of vegetative ramets produced is thought to vary stochastically, dependent on the individual’s characteristics, such as root crown diameter, which is recorded each year. Each new ramet becomes an individual to be followed up in its own right, appearing in the population in the year following its production. In its flowering year, a number of flowering heads and seeds may be produced in addition to vegetative ramets. A proportion of the seeds successfully establish and become ramets to be followed up in their own right, also appearing in the population in the year following their production. Seedlings and new ramets are distinguishable, but the number of each that a ramet generates in any given year is not observed directly - only the total number of ramets and the total number of seedlings are observed. However, in some cases a new vegetative ramet appears in physical proximity to where a ramet flowered or died the previous year. It is then reasonable to assume that the

new ramet was produced by the deceased ramet and this information can be incorporated into the life-cycle model.

This paper studies a large population of wavyleaf thistle located at seven sites within a sand prairie nature reserve in midwestern USA (Louda & Potvin, 1995). The data studied comprise a set of longitudinal data on size and status of 3320 above ground rosettes taken across 12 years. A summary of the data notation used is presented in the Appendix.

1.2. Bayesian models in demography

Bayesian approaches have been used to model missing information and individual variation based on longitudinal data from natural populations in ecology (Morgan, 2000; Cam *et al.*, 2002). In particular, there has been considerable work on techniques for estimating survival parameters from capture-recapture data of bird populations (Brooks *et al.*, 2002). Demographic modelling has been more rarely directly allied to Bayesian techniques for forecasting, although Clark (2003) and Clark *et al.* (2005) developed hierarchical models for capture-recapture data encompassing estimation error, variability among individuals, and discrete population structure and used them to predict population level outcomes. The propagation of stochastic elements into population level forecasts improves realism of predictions (Clark, 2003). A ubiquitous problem in demographic modelling is the presence of unobservable life-stages. Mark-recapture approaches to model this have been developed (Kery *et al.*, 2005). However, to our knowledge, there have been no integrated models modelling both such life stages and their effects on other demographic transitions. Furthermore, Bayesian population models for forecasting where populations are structured along a continuous variable (rather than discrete categories, Ellner & Rees, 2005) have not yet been developed.

2. Probability Models and Likelihood

The observed data comprise life history information, observations of plant size over the life history (Fig. 1), and counts related to reproduction and flowering; and the resulting likelihood has discrete and continuous components. In addition, there is a missing data component, relating to important but unobserved values (ramet vegetative reproduction and interconnection). Thus, overall, the likelihood and resulting posterior are complex.

Specifically, there are four different components of the model to specify; Stochastic models are required for:

- (a) the lifetime distribution of the ramet, that is the number of years until the terminal event, either death or flowering
- (b) the growth in ramet size
- (c) the number of flowering heads produced by a ramet
- (d) the number of vegetative ramets produced by a ramet

In the following model description, we will make parametric and distributional assumptions based on the observed data. We note here that the specific assumptions made may be relaxed or varied without significant increase in inferential complexity, as the proposed computational solution is quite general.

2.1. Missing ramet data

The growth in size and number of flowering heads are likely to depend on the number of ramets produced, because the resources available to each individual ramet are limited and must be

allocated to either action. Gaining understanding of this trade-off is essential to understanding the evolution of reproductive strategies. One complication is that density dependence appears to operate in this system, that is, the total number of individual ramets present at a given geographical location directly affects the establishment rate of new vegetative ramets. This leads to a discrepancy between the number of ramets observed in the dataset, and the number that were attempted by the ramets present. It is the latter that will affect growth in size and number of flowering heads; we therefore develop and describe an appropriate model to estimate this component. We present probability models for each of the four components needed to specify the model below, and then the full likelihood model.

2.2. The lifetime distribution

To construct a general model for this situation, a discrete lifetime distribution is considered; this will be denoted by $q = (q_1, q_2, \dots)$, where $q_a = P[\text{Terminal event at year } a]$ for $a = 1, 2, \dots$. The survivor function S_a is defined by

$$S_a = P[X > a] = 1 - q_1 - q_2 - \dots - q_a \quad a = 1, 2, \dots$$

and discrete hazards h_a are defined by the following conditional probability

$$h_a = P[\text{Terminal event at } a | \text{Survival until at least } a] = \frac{P[X = a]}{P[X \geq a]} = \frac{q_a}{S_{a-1}} \quad a = 1, 2, \dots,$$

The elementary relationship between these quantities is then

$$q_a = h_a \prod_{j=1}^{a-1} (1 - h_j), \quad S_a = \prod_{j=1}^a (1 - h_j) \quad a = 1, 2, \dots$$

The likelihood contribution for a ramet that dies/flowers in year a of its lifetime can thus be written as

$$(1 - h_1)(1 - h_2) \dots (1 - h_{a-1}) h_a$$

whereas if the lifetime is censored, the likelihood contribution is

$$(1 - h_1)(1 - h_2) \dots (1 - h_a).$$

If it is thought that the hazard probabilities are dependent on other available data, a discrete time proportional odds model can be adopted, that is

$$\frac{h_{ia}}{1 - h_{ia}} = \exp \{ \beta_{h1} + \beta_{h2} \log y_{ia} \} \frac{h_a}{1 - h_a}$$

where $\{h_a\}$ represents a set of baseline hazard probabilities, and the β_h parameters determine how much the baseline age specific hazard is altered by an individual i 's size at age a , y_{ia} . The proportional odds model used in this discrete survival model is the equivalent of the proportional hazards model used in the analysis of continuous time survival data.

This basic lifetime model needs to be modified to account for the fact that the terminal event can be either death or flowering, that is, that there is a competing risk element. Let (h_a, g_a) denote the probabilities of death and flowering at lifetime length a respectively. For a ramet that dies or flowers at time a , the likelihood contribution in these two cases is

$$h_a \prod_{j=1}^{a-1} (1 - h_j - g_j) \quad \text{or} \quad g_a \prod_{j=1}^{a-1} (1 - h_j - g_j)$$

respectively, whereas censored observations with censoring at a contribute

$$\prod_{j=1}^a (1 - h_j - g_j).$$

The proportional odds modification can be introduced in the same way as indicated above

$$\begin{aligned} \frac{h_{ia}}{1 - h_{ia} - g_{ia}} &= \exp \{ \beta_{h1} + \beta_{h2} \log y_{ia} \} \frac{h_a}{1 - h_a - g_a} = \exp \{ \beta_{h1} + \beta_{h2} \log y_{ia} \} \omega_a \\ \frac{g_{ia}}{1 - h_{ia} - g_{ia}} &= \exp \{ \beta_{g1} + \beta_{g2} \log y_{ia} \} \frac{g_a}{1 - h_a - g_a} = \exp \{ \beta_{g1} + \beta_{g2} \log y_{ia} \} \varpi_a \end{aligned}$$

2.3. The growth of the ramet

The growth of a ramet over its lifetime can be modelled using a linear Gaussian autoregression on the log scale, that is,

$$\log x = \gamma_0 + \gamma_1 \log y + \varepsilon$$

where $\log x$ and $\log y$ are measures of ramet size in successive years, and $\{\varepsilon\}$ are independent normal random errors (Fig. 1a). This model may be extended to include further independent variables, such as number of ramets produced (referred to as N , see Appendix), i.e., for individual i ,

$$\log x_i = \gamma_0 + \gamma_1 \log y_i + c_1 N_i + \varepsilon$$

where N_i is the number of vegetative ramets produced by individual i and the parameter c_1 describes the degree to which it affects growth. Production of vegetative ramets might be expected to have a negative impact on growth as less resources will be available for each ramet. This will have critical implications for the evolution of mixed reproductive strategies.

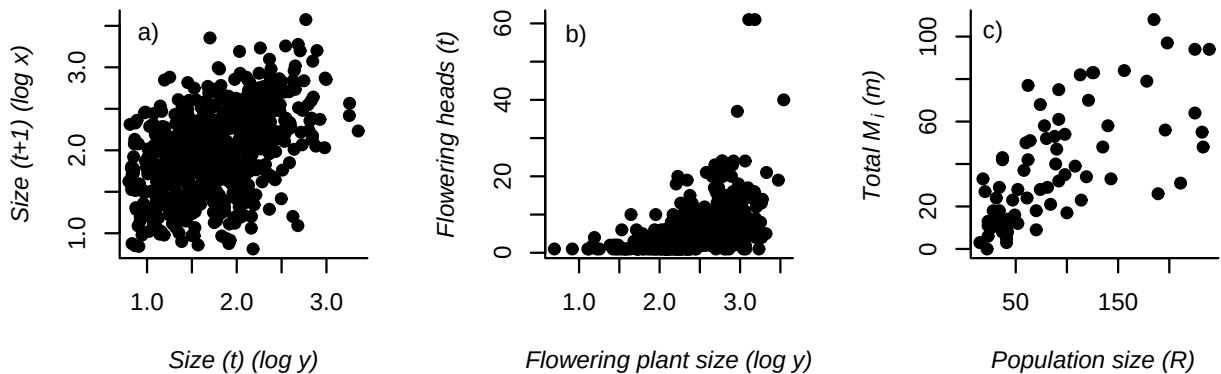


Fig. 1. a) Growth in size data b) Flowering head data c) Relationship between total population size, R , and observed total ramet numbers, m suggesting density dependent effects as the rate of increase in m declines with increasing R

2.4. The number of flowering heads and seedlings

To construct a model for the number of flower heads, which is size dependent and always greater than zero, we chose a Poisson model restricted to be strictly positive, with the mean number of flower heads produced by an individual ramet given by

$$\eta_0 + \eta_1 \log y$$

as in Fig. 1b. Again, it is possible to modify this model further by including further independent variables, such as number of ramets produced; so, for each individual we have

$$\eta_0 + \eta_1 \log y_i + c_2 N_i$$

where the parameter c_2 indicates the degree to which production of a vegetative ramet reduces the number of flower heads produced. Following a similar argument to that above, this parameter is expected to be negative. A constrained negative binomial model was also considered, but the amount of overdispersion observed in the sample was minimal.

The total number of seedling recruits the following year, directly observed in the data, b , is presumed to be the outcome of a binomial process on the seeds produced by flowering heads, i.e. $b \sim \text{Binomial}(S, p_e)$ where p_e is the probability that a seed successfully establishes. For simplicity, since no direct data is available on seed set, and seed production is expected to be proportional to flowering head number, we take $S_{j+1} = \sum_{i=1}^{F_j} (t_i + 100)$ where F_j is the total number of flowering ramets in year j , and t_i the observed number of flowering heads for each individual, to ensure $S \geq b$ for all years.

2.5. The number of ramets

The most appropriate model for the number of ramets produced vegetatively in any year (referred to here as N , see Appendix) is Poisson distributed, with rate parameter dependent on size, so that $N \sim \text{Poisson}(\lambda)$ where,

$$\log \lambda = \theta_1 + \theta_2 \log y$$

Although the number of ramets produced by each individual is not directly observed, we do observe a total number of vegetative ramets appearing the following year. The data suggest the action of density dependent processes, i.e., the observed total number of ramets is dependent on the total number of individuals present in the population, R (Fig. 1c). This indicates that the total number of ramets observed in each year, m , must be less than the number that were attempted for that year, N . We presume that this occurs according to a binomial probability model, specifically, $M \sim \text{Binomial}(N, \pi)$, where π is the probability of success of attempted vegetative ramets, and is a decreasing function of the total number of individuals present at each site at each time step, R . We chose π such that

$$\pi(\delta, R) = 1/(1 + |\delta|R).$$

We do not have perfect observation of the vegetative ramet production process. Our observed data comprise two pieces of information; the total number of ramets that were newly produced that survive into the following year at a given site in a given year, and an indicator variable, V , for each ramet reflecting whether that ramet was observed to produce vegetative ramets ($V = 1$) or not ($V = 0$) in any year. We have $V = 1$ for 1153 observations out of 6981, but there is no information on whether ramets produced new ramets or not for the remainder. Note that even when $V = 1$, the total number of ramets produced by the individual ramet is not observed in any instance.

2.6. Likelihood construction

Below, we give the derivation of the full-data likelihood after completion using Markov chain Monte Carlo to impute the missing values. Note that, even if the details of the models described above are varied (for example by replacing Poisson assumptions by negative binomial, or including or excluding variables from the regression model linear predictors), the same general computational framework can be used.

2.6.1. The Full Likelihood

The full likelihood is given by

$$L_1(X; y, \gamma, \sigma, Y, N) \times L_2(T; y, \eta, Y, N) \times L_3(N; y, \theta) \times L_4(M; N, \delta) \times L_5(z; y, \beta) \times L_6(S; t, p_e, b)$$

L_1 corresponds to the growth model outlined in Section 2.2, L_2 corresponds to the flowering heads model outlined in Section 2.3, L_3 and L_4 are models for ramets produced and establishing based around models introduced in Section 2.4, L_5 is the lifetime distribution model outlined in Section 2.1 and L_6 is the seed establishment model outlined in Section 2.3. Note that the number of ramets attempted, N , appears in L_1 and L_2 . Notation details are provided in the Appendix. Given knowledge of N , it is straightforward to simulate from L_1 , L_2 , L_5 and L_6 . Obtaining N by simulation from L_3 and L_4 is more complex. We address this in the next section.

2.6.2. The Culling Model

Consider the ramet production model (relating to L_3 and L_4) in more detail. For any given year, for the number of ramets attempted, N , and that succeed M , we have a likelihood that is the product of a Poisson, a binomial and a multivariate hypergeometric distribution. The Poisson distribution relates to ramet production such that:

$$\prod_{i=1}^R \frac{\lambda_i^{N_i} \exp\{-\lambda_i\}}{N_i!}$$

where R is the total number of individual above-ground rosettes in a given site and year. The binomial is the outcome of the fact that only a certain proportion of the ramets produced succeed in every year, such that:

$$\binom{N}{M} \pi^M (1 - \pi)^{N-M}$$

where $N = \sum_{i=1}^R N_i$ and $M = \sum_{i=1}^R M_i$ is directly observed as m , the number of successful new ramets observed the following year; π is included as a further parameter of interest, indicating the operation of density dependent processes on successful establishment of attempted vegetative ramets. The multivariate hypergeometric reflects that the binomial process operating at the yearly level occurs for a different total N at the level of each individual ramet in each site's population in each year, such that:

$$\frac{\prod_{i=1}^R \binom{N_i}{M_i}}{\binom{N}{M}}$$

Taking the product of these three functions, we are left with a distribution proportional to

$$\left\{ \prod_{i=1}^R \frac{\lambda_i^{N_i} \exp \{-\lambda_i\}}{N_i!} \right\} \times \left\{ \pi^M (1 - \pi)^{N-M} \right\} \times \left\{ \frac{\prod_{i=1}^R \binom{N_i}{M_i}}{\binom{N}{M}} \right\} \quad (1)$$

for each year of the data in each site. This is a convenient representation, as we observe $M = m$ in each case, for each site and in each year. Hence, when we sample the $M_1, \dots, M_R$ for any year conditional on $N_1, \dots, N_R$, we can propose any numbers that sum to m , such that $V_i \leq M_i \leq N_i$.

2.7. Sampling the Unobserved Ramet Counts

We begin by considering how to sample the unobserved ramet counts (M_i, N_i) for $i = 1, \dots, R$ from their full conditional posterior distributions.

2.7.1. Sampling the N_i given the M_i

For each N_i , from (1) we have that the full conditional for N_i is proportional to

$$L_1 L_2 \times \frac{\lambda_i^{N_i} \exp \{-\lambda_i\}}{N_i!} \times (1 - \pi)^{N_i} \times \frac{N_i!}{(N_i - M_i)!}$$

which is proportional to $L_1 L_2 \times g_N(N_i)$ where

$$g_N(N_i) = \frac{\{\lambda_i (1 - \pi)\}^{N_i}}{(N_i - M_i)!} \propto \frac{\{\lambda_i (1 - \pi)\}^{N_i - M_i}}{(N_i - M_i)!} \quad N_i \geq M_i.$$

The function g_N defines a probability distribution for N_i on $\{M_i, M_i + 1, \dots\}$ that can be sampled directly, as

$$N_i^{new} \stackrel{\mathcal{L}}{=} \text{Poisson}(\lambda_i (1 - \pi)) + M_i$$

Sampling from this model as a proposal for N_i leaves the MH acceptance probability term as

$$\min \left\{ 1, \frac{L_1(N_i^{new}) L_2(N_i^{new})}{L_1(N_i^{old}) L_2(N_i^{old})} \right\}$$

2.7.2. Sampling the M_i given the N_i .

From (1) we have that the full conditional for $(M_1, \dots, M_R)$ given $(N_1, \dots, N_R)$, and given that $\sum_{i=1}^R M_i = m$, is equal to

$$\frac{\binom{N_1}{M_1} \binom{N_2}{M_2} \dots \binom{N_R}{M_R}}{\binom{N}{M}}$$

which, it initially seems, we may sample directly as a multivariate hypergeometric distribution. However, such a direct approach does not incorporate the constraints on M_i imposed by the observed data. Re-writing the relevant part of the likelihood in equation (1) in a more basic form, we have

$$\prod_{i=1}^R \left\{ \binom{N_i}{M_i} \pi^{M_i} (1 - \pi)^{N_i - M_i} \right\} \quad (2)$$

The challenge is to sample $M_i, i = 1, \dots, R$ subject to the model and the constraints; in particular, for a subset of ramets, where $V_i = 1$, we have $M_i > 0$. Rather than using the full multivariate hypergeometric model above, we must sample the M_i in turn from their full conditionals. To retain the constraint $\sum_{i=1}^R M_i = m$, a pair of M_i values is chosen at each update, M_{i1} and M_{i2} . M_{i1} can take any value x in the range $x_l \leq x \leq x_u$ defined by the constraints that the sum $M_{i1} + M_{i2}$ must remain constant, $M_i \leq N_i$, and $V_i \leq M_i$. This also defines the range that M_{i2} can cover. Up to proportionality, we know that

$$p(M_{i1} = x | x_l, x_u, \pi) \propto \binom{x_u}{x} \pi^x (1 - \pi)^{x_u - x} \binom{x_u}{x_u - x} \pi^{x_u - x} (1 - \pi)^x \quad x = x_l, \dots, x_u$$

which defines a discrete distribution on a finite range which may be sampled directly.

2.8. Analysis of simulated data: Robustness issues

We used a simplified simulation to explore the behaviour of the model in order to assess the impact of different proportions of missing (that is, unobserved) ramet count data. We created a sample population of ramets, with a starting number of individuals chosen to reflect the data (~ 50 ramets for seven sites). Initial sizes for each individual were obtained as deviates from a normal distribution with mean and variance given by the mean and variance in log size of individuals of age 1 in the data ($\mu_{age_1} = 1.49, \sigma_{age_1}^2 = 0.35$). The fate of each of these simulated ramets (survival, death, flowering) was generated using chosen parameters, and the equations given in Section 2.1. Likewise, total numbers of ramets attempted by each individual (Section 2.4), number of flowering heads (Section 2.3) for those ramets whose fate was to flower, and new size for those ramets whose fate it was to neither die nor flower (Section 2.2) were generated. The simulation was then initiated. All computation was performed using the statistical language R.

At each time step, new individuals appear into the population: either as successful seedlings from seeds produced the previous year and establishing according to the chosen parameter value for probability of seed establishment p_e (Section 2.3); or as ramets produced from ramets present the previous year, which successfully establish according to the current total number of ramets in the population and the chosen value of the parameter δ (Section 2.4). Starting sizes are generated for these according to μ_{age_1} and $\sigma_{age_1}^2$ as above. The fate of these individuals and individuals surviving from the previous year is then established, based on their size and chosen parameter values (Section 2.1); and, the process is repeated. We simulated twelve years across seven sites to match the data available, obtaining a total of ~ 10000 observations across ~ 3000 individuals. We then applied the model described above to these simulated data to explore the ability of our model to make inference from this type of data.

Since inference on missing values denoting vegetative reproduction was a particular focus of this study, we tested our model assuming that different proportions of the N_i and M_i observations were completely known. In the corresponding MCMC simulations, the proportion of individual M_i and N_i values assumed to be known contributed directly to estimating the

likelihood of candidate parameters; the remaining M_i and N_i assumed to be missing were simulated from using the approaches outlined in Section 2.6. We then compared estimates of parameter values obtained by these MCMC simulations with parameter values corresponding to the maximum-likelihood estimates for the whole dataset, taking all missing information on M_i and N_i as known.

Results from the simulations indicate that although the accuracy (in terms of posterior variance) of the Bayesian estimate decreases when less information is available, there is little evidence of significant bias, that is, the true value often falls within the 95 % posterior credible interval even if no information is available on N and M beyond the constraints available in the data (see Fig. 2). This result is encouraging, and further confirms the advantages of the full Bayesian approach.

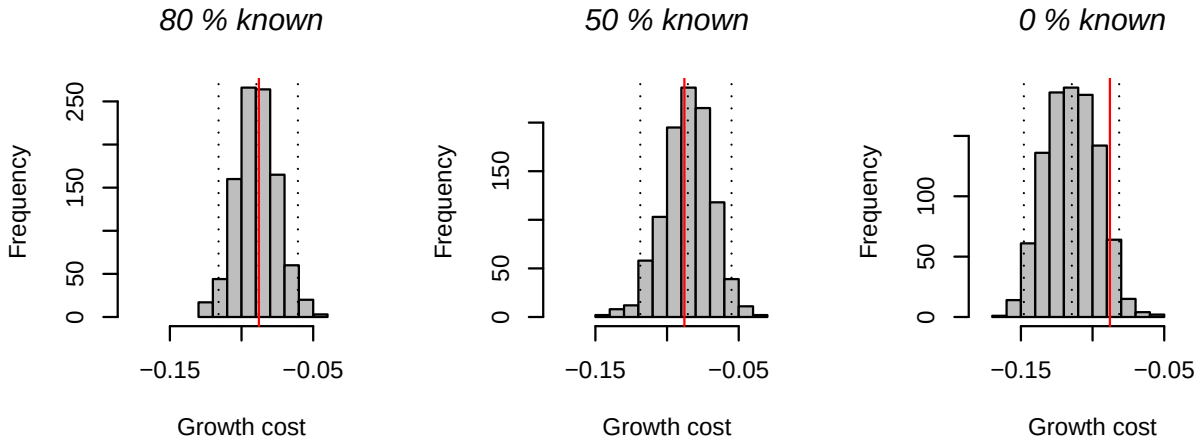


Fig. 2. Posterior distributions on the c_1 parameter from the model for growth in size from simulated data with different percentages of known information for the M_i and N_i , taken over 1000 iterations after a burnin of 1000 iterations. A solid vertical line indicates the maximum likelihood cost parameter value estimated by taking all M_i and N_i as known. Dotted lines indicate quantiles of the posterior distributions.

3. Analysis: Real Data

Preliminary exploration suggested that site-specific intercepts should be fitted to the growth, flower heads, and size dependent hazards functions; site specific intercepts were also fitted in the ramet production function. The slope of ramet production and flowering heads were sampled on $[0, +\infty)$, as it is unlikely that these parameters would be negative. A burn-in of 10000 was used, with thinning every 100 iterations. Mean and standard deviation of the posterior distribution for parameters for the first site obtained for the run which converged at the highest likelihood values (mean log likelihood ~ -24400) across 5000 iterations are presented in Table 2. Time series plots and posteriors for four key parameters are shown in Figure 3. There is evidence that production of a ramet reduces the number of flower heads produced in the following year (the 95 % posterior credible interval for c_2 is entirely in the negative domain), but increases ramet growth in size (the 95 % posterior credible interval for c_1 is entirely in the positive domain). There is no strong evidence for a size parameter in the ramet production function, that is, the the posterior distribution of θ_2 is concentrated at zero.

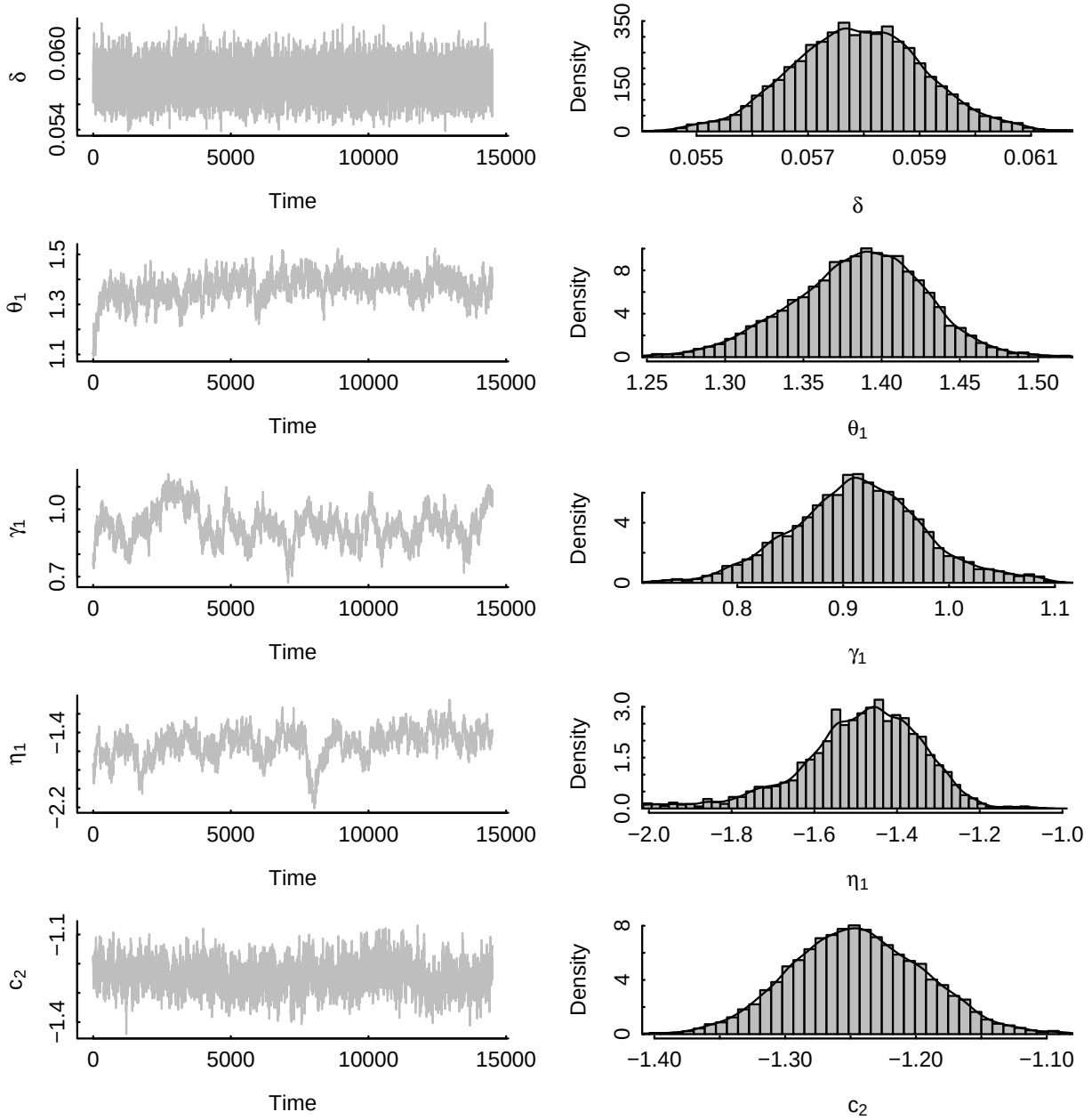


Fig. 3. Time series plots monitoring the MCMC output with histograms and kernel density estimates of 5000 samples after a burn-in of 10000 for 4 key parameters: the overall density dependence parameter δ , and for the first site, the intercept of vegetative ramet production, θ_1 , growth γ_1 , and flowering head production η_1 , and finally the overall cost parameter c_2 of vegetative ramet production on flowering head production. Note that the posterior of this parameter is entirely below zero.

Table 2. Parameter values obtained for the first site.

Parameter description		Mean (SD)
Density parameter	δ	0.057 (0.001)
Intercept ramet function	θ_1	1.381 (0.040)
Slope ramet function	θ_2	0.016 (0.013)
Intercept growth function	γ_1	0.921 (0.066)
Slope growth function	γ_2	0.446 (0.014)
Variance in growth	σ	0.200 (0.005)
Ramet cost on growth	c_1	0.193 (0.062)
Intercept head function	η_0	-1.445 (0.121)
Slope head function	η_1	1.522 (0.042)
Ramet cost on heads	c_2	-1.240 (0.052)
Baseline mortality hazard	h_0	0.421 (0.038)
Baseline flowering hazard	g_0	0.006 (0.002)
Size mortality intercept	β_{h1}	-0.000 (0.147)
Size mortality slope	β_{h2}	-0.144 (0.052)
Size flowering intercept	β_{g1}	-4.543 (0.227)
Size flowering slope	β_{g2}	3.381 (0.132)
Seedling establishment prob	p_e	0.052 (0.006)

4. Conclusions

The statistical model we describe here incorporates all important demographic aspects of this species' life history and it allows linkage between different demographic rates to be explicitly considered. Species' demographic rates are necessarily integrated, as in most systems resources will be limited, so this will be critical for understanding ecological and evolutionary outcomes. Most interestingly, although the results provide support for a negative effect of vegetative ramet production on flowering head production ($c_2 < 0$), the effect on ramet growth is positive ($c_1 > 0$). Unexplained heterogeneity across individuals might explain this pattern. Some individuals might be in spatial areas of high resource availability, allowing both increased growth and vegetative ramet production. This could be explored by including individual effects through a mixed modelling approach.

The detection of trade-offs between sexual and asexual reproduction (i.e., $c_2 < 0$) has implications for evolutionary outcomes as well as population dynamics and spread. Conveniently, the statistical framework is an exact reflection of the theoretic demographic model that could be used to predict population outcomes, the Integral Projection Model, a key tool for ecological forecasting and evolutionary developments (Ellner & Rees, 2005), facilitating the transition from parametrization to forecasting. Further possible extensions also suggest themselves. For example, parameter estimates across the seven sites suggest that high mortality is linked with high growth; and such an additional connection between demographic traits also could be incorporated in this model following approaches such as that used by Cam *et al.*, 2002.

A key innovation of the model is its direct incorporation both of missing stages, and how they are affected by population level processes. Understanding the operation of density dependence is one of the key challenges in population modelling (Lande *et al.*, 2006). The framework we present is generally applicable to systems with missing information on life stages where density dependence is thought to be operating, and extensions such as those suggested above are readily incorporated. This approach therefore may be of considerable use in developing novel inference fundamental to addressing a range of ecological or evolutionary questions.

Appendix

Notation

- **Data:** for each ramet i in each year
 - the size y_i and size the following year x_i
 - the survival status z_i (death, flower, nothing)
 - the number of flowering heads t_i if the ramet flowered
- **Data:** for each year j at each site
 - the total number of recruits from seed b_j
 - the total number of recruits from vegetative ramets m_j
- **Missing Data:** for each ramet i in each year
 - the number of ramets attempted N_i
 - the number of ramets produced M_i
- **The constraints:**
 - the total number of ramets produced at a site in any year is observed, that is, we observe

$$\sum_{i=1}^R M_i = m$$

where R is the number of ramets at the site in the current year. Furthermore, we observe, for some ramets, the fact that the ramet did produce ramets; that is, for some i , we observe $M_i > 0$, or equivalently $V_i = 1$.

The model and likelihood components

- **Growth in size is modeled by:**

$$\log X_i | y_i, \gamma, \sigma, N_i \sim N(\gamma_0 + \gamma_1 \log y_i + c_1 N_i, \sigma^2)$$

Therefore, across the 12 years for each of the seven sites, the first part of the likelihood $L_1(X; y, \gamma, \sigma, N)$ is given by

$$\prod_{j=1}^{11} \left\{ \prod_{i=1}^{A_j} f_1(X_{ji}; y_{ji}, \gamma, c_1, \sigma, N) \right\}$$

with f_1 the normal density, A_j the total number of individuals that survive in each year j , and y_{ji} is a measure of ramet size. In the final year of the study, no data is available for sizes the following year, so $j < 12$.

- **Flowering heads are modeled by:**

$$T_i|y_i, \eta, N_i \sim \text{Poisson}(\eta_0 + \eta_1 \log y_i + c_2 N_i)$$

As above, $L_2(T; y, \eta, Y, N)$ is therefore

$$\prod_{j=1}^{12} \left\{ \prod_{i=1}^{F_j} f_2(T_{ji}; y_{ji}, \eta, c_2, N,) \right\}$$

with f_2 the poisson mass function, F_j the total number of flowering ramets in each year j , and y_{ji} is a measure of ramet size.

- **Ramets Attempted are modelled by:**

$$N_i|y_i, \theta \sim \text{Poisson}(\lambda_i)$$

where $\lambda_i = \exp\{\theta_0 + \theta_1 \log y_i\}$. $L_3(N; y, \theta)$ is given by

$$\prod_{j=1}^{11} \left\{ \prod_{i=1}^{R_j} f_3(N_{ji}; y_{ji}, \theta) \right\}$$

with f_3 the Poisson mass function, R_j the total number of ramets in each year j , and y_{ji} is a measure of ramet size. In the final year of the study, no data is available for total number of ramets attempted and successfully establishing the following year, so $j < 12$.

- **Ramets Observed are modelled by:**

$$M|N, \delta \sim \text{Binomial}(N, \pi(\delta, R))$$

L_4 is given by

$$\prod_{j=1}^{11} \left\{ \prod_{i=1}^{R_j} f_4(M_{ji}; N_{ji}, \delta) \right\}$$

where f_4 is deduced from the binomial model and the parameter δ determines the decreasing probability of ramet establishment with increasing R .

- **The lifetime distribution:** $z_i|y_i, \beta_h, \beta_g, h_0, g_0$ has baseline hazard matrix

$$\mathbf{h}_0 = (h_0, g_0, 1 - h_0 - g_0)$$

modified in a proportional odds model by the log size at the previous time point, $\log y_i$. There is rarely age dependence in such species, so the baseline hazard is presumed to be the same for all ages. $L_5(z; y, \beta)$ is therefore given by

$$\prod_{j=1}^{12} \left\{ \prod_{i=1}^{R_j} f_5(z_{ji}; y_{ji}, \beta_h, \beta_g, h_0, g_0) \right\}$$

where f_5 is a product multinomial with likelihood contribution dictated by the indicator z_{ji} and likelihood contributions described in Section 4.1. Candidates for h_0 and g_0 were proposed from a Dirichlet and the acceptance probability was altered accordingly. After preliminary explorations indicated the range of parameter values, we fitted a prior on the Dirichlet distribution to accelerate convergence.

- Seed establishment is modelled by:

$$B_j|p_e, S_j \sim \text{Binomial}(S_j, p_e)$$

where b_j is directly observed as the number of seedling recruits in every year j for each site separately and $S_j = \sum_{i=1}^{F_j} S_i$ where S_i is the seed number produced by flowering individual i in year j , estimated as $S_i = t_i + 100$. $L_6(S; t, b, p_e)$ is given by

$$\prod_{j=1}^{11} f_6(B_j, S_j, p_e)$$

where f_6 is the binomial mass function. In the final year of the study, no data is available for total number of seedlings that successfully establish the following year, so $j < 12$.

REFERENCES

- BROOKS, S. P., Catchpole, E. A., Coulson, T., Lebreton, J.-D. & Morgan, B. J. T. (in preparation). Recent progress in using marked individuals to understand animal population biology. *Ecol. Lett.*
- BROOKS, S. P., Catchpole, E. A., Morgan, B. J. T. & Harris, M. P. (2002) Bayesian methods for analysing ringing data. *J. R. Statist. Soc. C* 29, 187 - 206.
- CAM, E., Link, W. A., Cooch, E. G., Monnat, J.-Y. & E. Danchin (2002) Individual co-variation in life history traits: seeing the trees despite the forest. *Am. Nat.* 159, 96 - 105.
- CLARK, J. S. (2003) Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology* 84, 1370 - 1381.
- CLARK, J. S., Ferraz, G. A., Ogue, N., Hays, H. & DiConstanzo, J. (2005) Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. *Ecology* 86, 2232 - 2244.
- ELLNER, S. P. & Rees, M. (2005) Integral projection models for species with complex demography. *Am. Nat.* 167, 410 - 428.
- KERY, M., Gregg, K. B. & Schaub, M. (2005) Demographic estimation methods for plants with unobservable life-states *Oikos* 108, 307 - 320.
- LANDE, R., Engen, S., Saether, B-E. & Coulson, T. (2005) Estimating Density Dependence from Time Series of Population Age Structure *Am. Nat.* 168, 76 - 87
- MORGAN, B. J. T. (2000) *Stochastic Modelling*, Arnold, London.
- GARDNER, S. N. & Mangel, M. (1999) Modeling investments in seeds, clonal offspring, and translocation in a clonal plant. *Ecology*, 80, 1202-1220.
- LOUDA, S. M. & Potvin, M. A. (1995) Effect of inflorescence-feeding insects in the demography and lifetime fitness of a native plant. *Ecology*. 76:229-245.

- RADOSEVICH, S. R., Holt, J. S. & Ghersa, C. (1997) *Weed Ecology. Implications for Management* (2nd ed) John Wiley & Sons, Inc, New York.
- REES, M., Condit, R., Crawley, M. J., Pacala, S. W. & Tilman, D. (2001) Long-term studies of vegetation dynamics. *Science*, 293, 650-655.
- ROSE, K. E., Rees, M. & Louda, S. M. (2005) Demographic and evolutionary impacts of native and invasive insect herbivores: a case study with Platte thistle, *Cirsium canescens*. *Ecology* 86:453-465.