Using Bayesian inference to understand the allocation of resources between sexual and asexual reproduction.

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 information and density dependence. We indicate how, to make inference on life history transitions involving both missing information and density dependence, the stochastic models underlying each component can be combined with each other and with priors to obtain expressions that can be directly sampled. 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.

E-mail: cjm29@duke.edu

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

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 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.*, (2002, 2004)), explicit linkage to a full demographic model is rare. Our model and analysis provide such linkage.

Inference from our model is performed using a Bayesian framework via Markov chain Monte Carlo (MCMC). The most challenging aspect of the MCMC analysis is the imputation of missing values that form critical parts of the model. For example, allocation of resources to vegetative reproduction cannot be quantified without the unobserved number of vegetative offspring being incorporated explicitly into the model. We adopt the conventional computational approach to Bayesian missing data problems, and extend the MCMC accordingly.

1.1.   *The wavyleaf thistle species and data*

The wavyleaf thistle has two modes of reproduction: i) vegetative, where offspring are produced through branching from a parents' underground root system and are therefore genetically identical to them, and ii) sexual, where offspring result from seeds which are dispersed from flowers on the parent, and then establish in the soil - these latter are referred to as seedlings, and can be distinguished from vegetatively produced offspring by the presence of cotelydons, a special leaf form. The visible, above-ground plants in the field site are referred to as *ramets*. Because of the existence of vegetative reproduction, a single genetic individual may consist of several ramets, which branch off the same underground taproot. However, it is difficult to identify genetic individuals as the roots connecting above-ground ramets are deep under the ground. Consequently, 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. Since demographic features are recorded at discrete intervals (yearly) it is possible for a ramet to initiate a new ramet, and then be recorded as dead the following year. 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 by using callipers to measure the diameter of the roots at a point taken directly above the ground. 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

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. Although plants die on flowering, the mortality risk that occurs independently of flowering is the main interest in this paper. Death and flowering are consequently modeled as competing hazards. If flowering occurs, death does not.

Year	1	2	3	4	5	6
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	-

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.

In this paper, we study 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 plant size (root crown diameter) and status (dead, flowering or living) of 3320 ramets taken across 12 years. A summary of the data notation used is presented in the Appendix A.1.

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). 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 unobservable life stages have been developed (Kery *et al.*, 2005). However there have been no integrated models for modelling both life stages and their effects on other demographic transitions. Furthermore, Bayesian population models that are matched to models for forecasting where populations are structured along a continuous variable (rather than discrete categories, Ellner & Rees, 2005) have not yet been developed.

1.3. Plan of paper

In Section 2, the probability model used is described in detail; in particular the specific components for plant life history, growth, reproduction and flowering are detailed. In addition, the relationship between the observed quantities (plant size, life events, number of ramets and flowering heads produced) and unobserved quantities (number of ramets attempted) is defined.

Prior distributions for the Bayesian analysis are described. In Section 3, details of the MCMC computational algorithm are given, and in Section 4, a simulation study outlined. In Section 5 the analysis of the wavyleaf data set is described, and in Section 6, the ecological and demographic implications of the findings are discussed.

2. Probability Models and Likelihood

The observed data comprise life history information, observations of plant size over the life history (Figure 2(a)), and counts related to reproduction and flowering (Figure 2(b)). In addition, there is a missing data component, relating to important but unobserved values (ramet vegetative reproduction and interconnection). The goal of the Bayesian inference is to sample from a joint posterior distribution defined by all these elements. 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. Several of the underlying conditional distribution functions can be described by discrete or continuous likelihoods, others are more complex. 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. Once distributional properties of all the separate components have been defined, the full expression is assessed to identify any possible simplifications. In Section 2.7, we describe complete-data likelihood and Bayesian prior components, and then in Section 3 describe how sampling from the resulting posterior distribution is achieved via Metropolis-Hastings.

We consider both *across-site* models (where some parameters that appear in the models are common across sites, and *site-specific* models, where the parameters that appear in the proportional odds model, the growth model and the model for flowering heads are site specific. This will allow us to assess whether the covariates have different influences at different sites. Further details are given in Section 5.

2.1. Missing ramet data

The growth in size and number of flowering heads is 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 (Figure 1). In exploratory analysis, the slope of a linear regression linking log total number of new vegetative ramets appearing at time t with number of ramets present at time t is significantly less than 1 (0.88 ± 0.100, $n = 82$). Consequently, at high densities of ramets at t, there is some evidence that the number of new vegetative ramets asymptotes. The most immediate
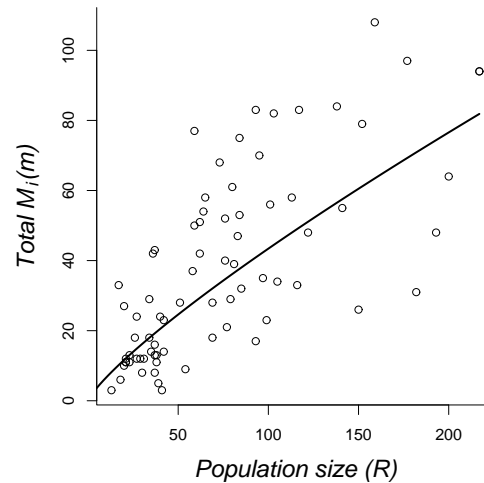
Fig. 1. Relationship between total population size, R, and observed total ramet numbers per site, m suggesting density dependent effects as the rate of increase in m declines with increasing R. The line indicates the fitted model from the linear regression taken on a log scale (see text).

ecological explanation is density dependence: i.e., although ramets allocate resources towards creating new vegetative ramets at equivalent rates at low and high densities, at high densities, these new offspring fail to successfully establish and die before the next population census. At high densities, resources such as space or nutrients are likely to be limiting, and the more vulnerable smaller plants are likely to suffer the consequences. Ramets that are produced but do not successfully establish are referred to as ramets "attempted".

Such density dependence will lead to a discrepancy between the number of ramets observed in the dataset, and the number that were initiated 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 derive the full (complete-data) likelihood.

In the remainder of the paper, although information from different sites is available, we use notation that suppresses this site-dependence; we index plants by i and year by t, and for convenience index by k when considering all combinations of indices i, t.

2.2. The lifetime distribution

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

$$S_a = P[X > a] = 1 - \sum_{j=1}^{a} q_j \qquad a = 1, 2, ...$$

and discrete hazards h_a are defined by the following conditional probability

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

The elementary relationship between these quantities is then

$$q_a = h_a \prod_{j=1}^{a-1} \left(1 - h_j\right), \qquad S_a = \prod_{j=1}^{a} \left(1 - h_j\right)$$

for each a. The likelihood contribution for a ramet finally observed at age a is thus

$$\begin{array}{rcll} q_a & = & \left(1 - h_1\right)\left(1 - h_2\right)\ldots\left(1 - h_{a-1}\right) h_a & \text{if the terminal event occurs} \\ S_a & = & \left(1 - h_1\right)\left(1 - h_2\right)\ldots\left(1 - h_{a-1}\right)\left(1 - h_a\right) & \text{if the lifetime is censored} \end{array}$$

In fact, the terminal event can be either death or flowering, so the model must be extended to allow for these competing risks. Let (h_a, g_a) denote the hazard probabilities of death and flowering at age a respectively. For a ramet that (i) dies, (ii) flowers or (iii) is censored at age a, the likelihood contributions are

$$\text{(i) } h_a \prod_{j=1}^{a-1} \left(1 - h_j - g_j\right) \qquad \text{(ii) } g_a \prod_{j=1}^{a-1} \left(1 - h_j - g_j\right) \qquad \text{(iii) } \prod_{j=1}^{a} \left(1 - h_j - g_j\right).$$

respectively. For the wavyleaf plants, there is rarely age-dependence in the hazard. That is, the baseline hazard is presumed to be the same for all ages, so that $h_a = h_0$ for all a. This assumption could readily relaxed for other species, however.

For plant i observed in the data in year t of the study, let $Z_{i,t}$ denote the random variable indicating the outcome for that plant in that year, that is, we observe $Z_{i,t} = z_{i,t}$ where
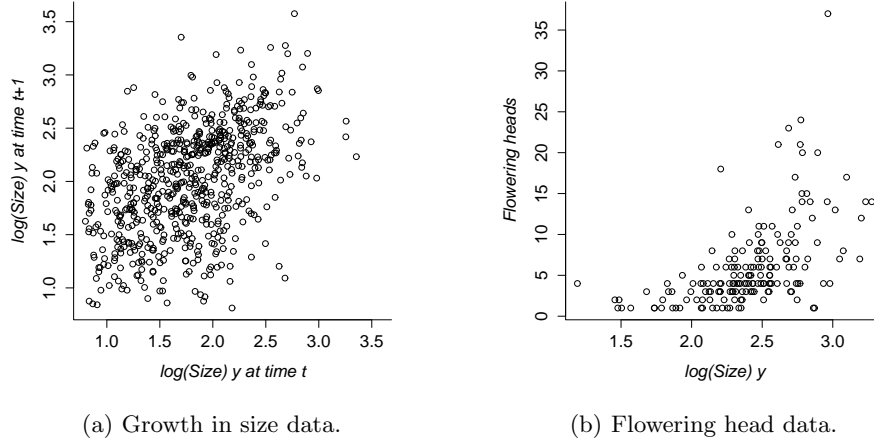
$$z_{i,t} = \begin{cases} 1 & \text{if the plant dies} \\ 2 & \text{if the plant flowers} \\ 0 & \text{if the plant survives} \end{cases}.$$

The likelihood contribution for plant i may be written by conditional independence as

$$\prod_t h_{i,t}^{\text{I}\{z_{i,t}=1\}} g_{i,t}^{\text{I}\{z_{i,t}=2\}} (1 - h_{i,t} - g_{i,t})^{\text{I}\{z_{i,t}=0\}} \tag{1}$$

where $h_{i,t}$ and $g_{i,t}$ are the plant-specific hazards at time t that incorporate information on the age of the plant in that year, $\text{I}\{.\}$ is the indicator function, and the index t extends over all years in which the plant is observed. The hazard probabilities can be made dependent on covariates, in a proportional odds model, where baseline hazards are modified using transformed linear predictors. Hence we have

$$\frac{h_{i,t}}{1 - h_{i,t} - g_{i,t}} = \exp\left\{\beta_{h1} + \beta_{h2} y_{i,t}\right\} \frac{h_0}{1 - h_0 - g_0} = \exp\left\{\beta_{h1} + \beta_{h2} y_{i,t}\right\} \omega_0$$

$$\frac{g_{i,t}}{1 - h_{i,t} - g_{i,t}} = \exp\left\{\beta_{g1} + \beta_{g2} y_{i,t}\right\} \frac{g_0}{1 - h_0 - g_0} = \exp\left\{\beta_{g1} + \beta_{g2} y_{i,t}\right\} \varpi_0$$

(a) Growth in size data.

(b) Flowering head data.

Fig. 2. Data for Site 2 for a) Ramet growth b) Ramet flowering head production.

where

$$\omega_0 = \frac{h_0}{1 - h_0 - g_0} \qquad\qquad \varpi_0 = \frac{g_0}{1 - h_0 - g_0}$$

are the baseline odds on death and flowering respectively, and β_{h1}, β_{h2} and β_{g1}, β_{g2} determine the dependence of the hazards on log size, $y_{i,t}$. Inverting the relationships, we obtain

$$h_{i,t} = \frac{\exp\left\{\beta_{h1} + \beta_{h2}y_{i,t}\right\}\omega_0}{1 + \exp\left\{\beta_{h1} + \beta_{h2}y_{i,t}\right\}\omega_0 + \exp\left\{\beta_{g1} + \beta_{g2}y_{i,t}\right\}\varpi_0}$$

and so on. If a single site is considered, parameters in the proportional odds model are only identifiable if the constraint $\beta_{h1} = \beta_{g1} = 0$ (or equivalent) is used. If variation across sites is to be considered, the model is only identifiable if we set $\beta_{h1} = \beta_{g1} = 0$ for some site - without loss of generality, we use the first. We have also experimented with a random effects or frailty specification across sites for this model component.

2.3. The growth of the ramet

The growth of a ramet over its lifetime will be modelled using a linear Gaussian autoregression on the log scale (Metcalf *et al.* 2003, see Figure 2(a)). We use the following model; if $y_{i,t} = \log x_{i,t}$ is the log-scale measure of ramet size, $N_{i,t}$ is the number of ramets produced by the same plant in the previous year, and $\{\varepsilon_{i,t}\}$ are independent normal random errors, then we have

$$y_{i,t+1} = \gamma_0 + \gamma_1 y_{i,t} + c_1 N_{i,t} + \varepsilon_{i,t} \tag{2}$$

where the parameter c_1 describes the degree to which ramet production affects growth. Production of vegetative ramets might be expected to have a negative impact on growth as fewer resources will be available for each ramet. This will have critical implications for the evolution of mixed reproductive strategies.

2.4.   The number of flowering heads

To construct a model for the number of flowering heads, $W_{i,t}$, produced by each flowering plant, which is size dependent and always greater than zero (Figure 2(b)), we model $W_{i,t}$ using a translated Poisson model, that is, $W_{i,t} - 1 \sim Poisson(\exp(\eta_{i,t}))$, where

$$\eta_{i,t} = \eta_0 + \eta_1 y_{i,t} + c_2 N_{i,t} \tag{3}$$

consistent with the exploratory analysis in Figure 1(b), but including the number of ramets produced at time t, $N_{i,t}$. 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 might be expected to be negative; resources are expended in the production of ramets.

A model for $W_{i,t}$ based on the zero-truncated Poisson distribution

$$P[W = w] = \frac{\lambda^w e^{-\lambda}}{(1 - e^{-\lambda})w!} = \frac{1}{w!}\exp\{w\log\lambda - \lambda - \log(1 - e^{-\lambda})\} \qquad w = 1, 2, \ldots \tag{4}$$

could also be used. The mean/variance relationship for the two models is different; if μ and $V(\mu)$ represent the mean and variance for the models, then for the translated Poisson model $V(\mu) = \mu - 1$, whereas for the truncated Poisson model,

$$\mu = \frac{\lambda}{1 - e^{-\lambda}} \qquad V(\mu) = \mu\left(1 - \mu e^{-\lambda}\right).$$

For large λ, the zero-truncated model reduces to the ordinary Poisson model with a linear (identity) mean-variance relationship, but for small λ, the mean-variance relationship is concave; we have to second-order for λ near zero

$$\mu \simeq 1 + \frac{\lambda}{2} + \frac{\lambda^2}{12} \qquad V(\mu) \simeq \frac{\lambda}{2} + \frac{\lambda^2}{6}$$

The zero-truncated Poisson is an Exponential Family distribution with canonical parameter $\eta = \log\lambda$, and thus a generalized linear model based on a linear predictor and log link for λ could be used, and the likelihood is easily modified to reflect this. However, interpretation of coefficients in the linear predictor is no longer so straightforward, due to the non-linear dependence of μ on λ. Despite this, we report results for both models.

2.5.   The number of seedlings

Data are also collected, on a site-by-year basis, on the total number of seedling recruits from the seeds produced by the flowering heads when a plant flowers. Unfortunately, the number of seeds produced is not observed on a plant-by-plant basis. We are therefore quite restricted in the model that we can use to model the seedling recruits. However, these data do have demographic importance in the ecological context, as we wish to infer the *establishment rate* of seeds produced. Our strategy is therefore as follows; in any site/year, the total number of seedling recruits the following year, B_t, which is directly observed in the data as b_t, is presumed to be the outcome of an independent binomial thinning process on the seeds produced by flowering heads, that is, conditional on S, the total number of seeds produced in any year, we have that $B_t \sim Binomial(S, p_e)$ where p_e is the probability that a seed successfully establishes. The quantity S is not observed, but rather than build a missing data model, we assume that

for each flowering head the number of seeds produced is a Poisson random variable with rate λ_S, independently across plants. As a consequence, we can deduce the model for the observed number of seedling recruits directly in terms of the observed number of flowering heads; in fact, for any year

$$B_t \sim Poisson(F\lambda_S p_e)$$

where F is the total number of flowering ramets at that site in the previous year. A range of other formulations is possible; this is the simplest and most realistic one. Seed production per head is unlikely to vary strongly between sites, however, the model can be easily extended to allow variation of the p_e quantity across sites, in which case parameter λ_S and the collection of p_e quantities can be estimated from the data. We restrict attention to the simpler model, in which the quantity $\lambda_{Se} = \lambda_S p_e$ is estimated from the available data, and represents the rate of seedling establishment per flowering head.

2.6.   The number of ramets

The number of ramets that are attempted to be produced vegetatively by plant i in year t, $N_{i,t}$, is unobserved and thus is a variable that is treated as missing data in the analysis. The model for the $N_{i,t}$ is effectively a prior distribution; the observed data are likely only to be minimally informative about the $N_{i,t}$, and so for our analysis we use a fixed $Poisson(\theta)$ prior, where θ is selected to reflect biological prior opinion; this is discussed in section 2.8. Extensions to this model are possible; for example, the rate parameter in this prior could be made dependent on size, if sufficient extra information was available.

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_t, at that site in that year (see Figure 1). This indicates that the total number of ramets observed in each year, $M_{i,t}$, must be less than the number that were attempted for that year, $N_{i,t}$. We presume that this occurs according to a binomial probability model, specifically, $M_{i,t}|N_{i,t} = n_{i,t} \sim Binomial\,(n_{i,t}, \pi_t)$, where π_t 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_t. We chose π_t such that
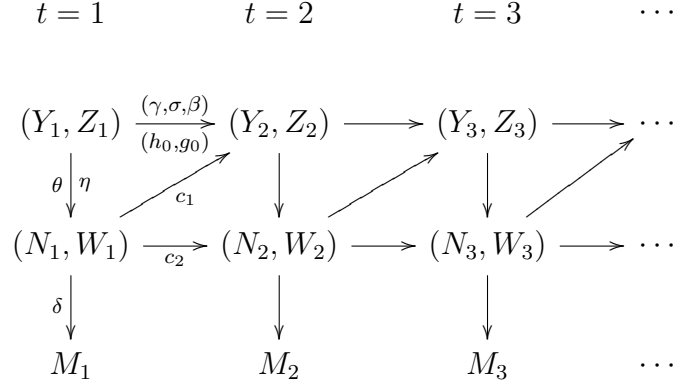
$$\pi_t = \pi(\delta, R_t) = 1/(1 + e^\delta R_t). \tag{5}$$

Finally, let

$$\widetilde{N}_t = \sum_{i=1}^{R_t} N_{i,t} \qquad \widetilde{M}_t = \sum_{i=1}^{R_t} M_{i,t}$$

denote the totals of ramets attempted and observed in each year. We do not have perfect observation of the vegetative ramet production process, but the observed data comprise two pieces of information; the total number, $\widetilde{M}_t$, of newly produced ramets that survive into the following year at a given site in a given year is observed to be equal to $\widetilde{m}_t$, and an indicator variable, $V_{i,t}$, for each ramet reflecting whether that ramet was observed to produce vegetative ramets ($V_{i,t} = 1$) or not ($V_{i,t} = 0$) in any year. Across all sites, we have $V_{i,t} = 1$ for 1153 observations out of 6981, but there is no such information available 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.

A diagrammatic summary of the inter-relationship between the various data sources and related parameters for an individual plant through time is given below. This diagram reflects the modelled conditional independencies and modelled links; with subscript i suppressed, recall that Y indicates a log size, and Z indicates survival status.

$$
\begin{array}{cccc}
t=1 & t=2 & t=3 & \cdots \\[2em]
(Y_1, Z_1) \xrightarrow[(h_0,g_0)]{(\gamma,\sigma,\beta)} (Y_2, Z_2) \longrightarrow (Y_3, Z_3) \longrightarrow \cdots \\[1.5em]
\theta \downarrow \eta \qquad \nearrow{c_1} \qquad \downarrow \qquad \nearrow \qquad \downarrow \qquad \nearrow \\[1em]
(N_1, W_1) \xrightarrow[c_2]{} (N_2, W_2) \longrightarrow (N_3, W_3) \longrightarrow \cdots \\[1.5em]
\delta \downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow \\[1em]
M_1 \qquad\qquad M_2 \qquad\qquad M_3 \qquad \cdots
\end{array}
$$

Information omitted from this diagram include site-level total numbers of plants attempted and recorded in each year $\widetilde{N}_t$ and $\widetilde{M}_t$, the number of seedling recruits b_t, and other constraints imposed by aspects of the observation process to be described earlier in this section.

2.7. Likelihood Construction and Prior Specification

2.7.1. The Complete Data Likelihood

Adopting the classical terminology of data augmentation problems, the full- or complete-data likelihood, containing the observed data vectors $\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{b}$ and the (augmenting, or missing) latent data vectors $\boldsymbol{n}, \boldsymbol{m}$, is given by

$$
\ell_1\left(\boldsymbol{y}; \gamma, \sigma, \boldsymbol{n}\right) \times \ell_2\left(\boldsymbol{w}; \boldsymbol{y}, \eta, \boldsymbol{n}\right) \times \ell_3\left(\boldsymbol{n}; \boldsymbol{y}, \theta\right) \times \ell_4\left(\boldsymbol{m}; \boldsymbol{n}, \delta\right) \times \ell_5\left(\boldsymbol{z}; \boldsymbol{y}, h_0, g_0, \beta\right) \times \ell_6\left(\boldsymbol{b}; \boldsymbol{w}, \lambda_{Se}\right) \quad (6)
$$

ℓ_1 corresponds to the growth model outlined in section 2.3, ℓ_2 corresponds to the flowering heads model outlined in section 2.4, ℓ_3 and ℓ_4 are models for ramets produced and establishing based around models introduced in section 2.6, ℓ_5 is the lifetime distribution model outlined in section 2.2 and ℓ_6 is the seed establishment model outlined in section 2.5. Note that the missing data, the number of ramets attempted, appears in ℓ_1 and ℓ_2, due to the presences of the costs-on-resources parameters c_1 and c_2.

2.7.2. The Culling Model

The conditional specifications described above are predicated on knowledge the missing ramet production data, $(M_{i,t}, N_{i,t})$, for all plants at all sites in all years. However, updating these parameters by simulating from their full conditional via ℓ_3 and ℓ_4 is not straightforward.

Consider the ramet production model at a single site in a single year, and the form of ℓ_3 and ℓ_4. For any given year t for plant i, for the number of ramets attempted, $N_{i,t}$, and the number that succeed $M_{i,t}$, we have a likelihood that is the product of a Poisson, a binomial and a multivariate hypergeometric distribution. We have the (prior) model for $N_{i,t}, i = 1, \ldots, R_t$ in

year t given by

$$\prod_{i=1}^{R_t} \frac{\lambda_{i,t}^{N_{i,t}} \exp\{-\lambda_{i,t}\}}{N_{i,t}!}$$

where, from Section 2.6, $\lambda_{i,t} = \theta$. Secondly, due to the culling model in section 2.7.2, we have a binomial model

$$\binom{\widetilde{N}_t}{\widetilde{M}_t} \pi_t^{\widetilde{M}_t} (1 - \pi_t)^{\widetilde{N}_t - \widetilde{M}_t}$$

where π_t is given by equation (5). Note that $\widetilde{M}_t$ is directly observed as $\widetilde{m}_t$, the number of successful new ramets observed the following year; recall that $\pi_t = \pi(\delta, R_t)$, and δ is included as a further parameter of interest. Finally, we have that

$$\binom{\widetilde{N}_t}{\widetilde{M}_t}^{-1} \prod_{i=1}^{R_t} \binom{N_{i,t}}{M_{i,t}} \tag{7}$$

that is, a multivariate hypergeometric distribution which reflects that the binomial process operating at the level of each individual ramet in each site's population each year occurs with respect to a total $\widetilde{N}_t$. Taking the product of these three functions, we are left with a distribution proportional to

$$\left\{ \prod_{i=1}^{R_t} \frac{\lambda_{i,t}^{N_{i,t}} \exp\{-\lambda_{i,t}\}}{N_{i,t}!} \right\} \times \left\{ \binom{\widetilde{N}_t}{\widetilde{M}_t} \pi_t^{\widetilde{M}_t} (1 - \pi_t)^{\widetilde{N}_t - \widetilde{M}_t} \right\} \times \left\{ \binom{\widetilde{N}_t}{\widetilde{M}_t}^{-1} \prod_{i=1}^{R_t} \binom{N_{i,t}}{M_{i,t}} \right\} \tag{8}$$

for each year of the data in each site. This is a convenient representation, as we observe the total $\widetilde{M}_t = \widetilde{m}_t$ for each year at each site, and hence, when we sample the $(M_{1,t}, ..., M_{R_t,t})$ for any year conditional on $(N_{1,t} = n_{1,t}, ..., N_{R_t,t} = n_{R_t,t})$, we can propose any numbers that sum to $\widetilde{m}_t$, such that $V_{i,t} \le M_{i,t} \le N_{i,t}$ for $i = 1, \ldots, R_t$. Note that $\widetilde{N}_t$ can be replaced by $\sum_{i=1}^{R_t} N_{i,t}$ in equation (8) whenever the $N_{i,t}$ values are being updated; see section 3.1.

2.8. Prior specification

To complete the Bayesian model specification, prior distributions should also be specified. In the analysis we present below, for the various linear predictor coefficient parameters γ, η, β and δ, and the cost parameters c_1 and c_2, priors were chosen to be relatively uninformative, aside from certain known positivity constraints (see section 2.5). The likelihood components for all these elements are straightforward, and a considerable amount of data pertaining directly to these parameters at the first stage of the hierarchy are available, thus we utilized proper but relatively diffuse prior distributions (independent Normal priors with variance 100) as a standard objective specification. We did perform rudimentary sensitivity analysis within this diffuse class of priors, but noticed little difference in inferences made. We note here that improper priors were not considered for these parameters, partly because of the possibility of impropriety in the resulting posterior distribution.

The prior for the autoregressive variance parameter σ^2 was, in contrast, chosen to be relatively informative. The size of ramets on the log-scale was thought, *a priori*, to vary in the range 0 to 3. The prior distribution for σ^2 was chosen to reflect this prior knowledge

of the marginal distribution of Y. Consequently, we selected an Inverse Gamma prior with parameters $\alpha_\sigma = 2.5, \beta_\sigma = 2$ and mean $2/(2.5\text{-}1)$; this gave a prior probability of approximately 0.93 that $0 < \sigma^2 < 3$. Sensitivity analysis with respect to this prior was also carried out, but again led to minimal changes in inference. Note that, here, a non-informative prior is acceptable - that is, does not lead to an improper posterior - but not particularly attractive as substantive, albeit generic, prior information is available.

The prior for the discrete baseline hazard probabilities in the lifetime distribution, (h_0, g_0), was chosen to be a Dirichlet$(1, 1, 1)$ distribution, uniform on the two dimensional simplex.

In the model for the unobserved ramet numbers $N_{i,t}$, the prior was chosen to reflect biological reality. With high probability, the number of ramets attempted by any plant is less than ten; indeed, a more realistic upper bound would be five. With this in mind, we began with a $Poisson(1)$ prior, and examined sensitivity of results when this prior was changed to be $Poisson(2)$ and $Poisson(5)$.

3.   Markov chain Monte Carlo Implementation

We now describe the Markov chain Monte Carlo (MCMC) strategy used. Generically, the simplest form of the Metropolis-Hastings algorithm for target posterior distribution Ψ, a product of the likelihoods and the prior distributions described above proceeds as follows. If the state of the chain $\{X_t\}$ at iteration t is given by $X_t = u$, then a candidate state v is generated from conditional density $q(u, v) = q(v|u)$, and accepted as the new state of the chain (that is, we set $X_{t+1} = v$) with probability $\alpha(u, v)$ given by

$$\alpha(u, v) = \min\left\{1, \frac{\Psi(v)q(v, u)}{\Psi(u)q(u, v)}\right\}.$$

The usual MCMC approach to missing data inference problems involves using a Gibbs sampler strategy, and we adopt this strategy here. Specifically, we sample the parameters conditional on the imputed missing values, and then sample the missing values conditional on the current values of the parameters. Given knowledge of $\{N_k : k \in \mathcal{K}\}$, where $\mathcal{K}$ is the set of all pairs of indices of plant and year, it is straightforward to simulate values from the conditional posterior distribution for system parameters $(\gamma, \eta, \beta, \delta, \sigma)$. In particular sampling is straightforward using the Metropolis-Hastings with Gibbs algorithm; the Markov chain is initialized, and then candidate values for subsets of the parameters are proposed and accepted or rejected iteratively in the usual way, conditional on fixed values of the remaining parameters, with the subsets being updated in turn. The most effective strategy that we found involved using the Metropolis algorithm with a symmetric proposal density $(q(u, v) = q(v, u))$; we chose to sample candidates from a normal distribution around the current parameter values, with variance chosen by tuning from pilot runs. In some cases, updating from the full conditional distribution of a parameter is possible; for example, the seed establishment rate quantity λ_{Se} has a Gamma full conditional density which can be sampled directly.

Sampling some of the missing data values, in particular the N_k, M_k values described in previous sections under the constraints of the model, in an efficient fashion is more problematic. We study the related problems in the following subsections.

3.1.   Sampling the N_k given that $M_k = m_k$

The full conditional posterior distributions for $N_k, k \in \mathcal{K}$ can be obtained directly from the complete-data likelihood in equation (6). We propose two methods for sampling these vari-

ates. In Method I, we use a conventional Metropolis-Hastings proposal based on a Poisson approximation. For Method II, we use truncation and exact sampling from the discrete full conditional distribution; details are given in Appendix A.4. Results indicate that Method I and Method II give identical posterior distributions for all the parameters, indicating that the truncation has no effect on the inference. In all the results below, we utilize Method II implemented for all plants in parallel, as the N_k are conditionally independent.

3.2. Sampling the M_i given the N_i.

¿From (8) we have that the full conditional distribution for $(M_{1,t}, ..., M_{R_t,t})$ given the number of ramets attempted $(N_{1,t}, \ldots, N_{R_t,t}) = (n_{1,t}, \ldots, n_{R_t,t})$, and given that the total $\widetilde{M}_t = \widetilde{m}_t$, is a multivariate discrete distribution proportional to the term in equation (7), that is, proportional to a multivariate hypergeometric distribution. However, a direct sampling approach using the multivariate hypergeometric does not respect the constraints on $M_{i,t}$ imposed by the observed data. Rewriting equation (8), we have the full conditional joint mass function for $(M_{1,t}, ..., M_{R_t,t})$

$$p(m_{1,t}, \ldots, m_{R_t,t}) = \prod_{i=1}^{R_t} \left\{ \binom{n_{i,t}}{m_{i,t}} \pi_t^{m_{i,t}} (1-\pi_t)^{n_{i,t}-m_{i,t}} \right\}. \tag{9}$$

The objective is to sample $M_{i,t}, i = 1, \ldots, R_t$ subject to the model and the constraints; in particular for a subset of plants, where $V_{i,t} = 1$, we have $M_{i,t} > 0$. Rather than update using the full conditional joint mass function above, we sample the $M_{i,t}$ in turn from their full conditionals. To respect the constraint $\widetilde{M}_t = \widetilde{m}_t$, a pair of indices (i_1, i_2) is chosen uniformly at each iteration, we then update $M_{i_1,t}$ and $M_{i_2,t}$ from their joint full conditional distribution, given values of all other $M_{i,t}$ fixed at $m_{i,t}$. Now, $M_{i_1,t}$ can take any value x in the range $x_l \leq x \leq x_u$ defined by the constraints

(i) that the sum $M_{i_1,t} + M_{i_2,t} = m_{i_1,t} + m_{i_2,t} = m_{i_1,i_2}$ must remain fixed during the update

(ii) $M_{i_1,t} \leq n_{i_1,t}$ and $M_{i_2,t} \leq n_{i_2,t}$

(iii) $V_{i_1,t} \leq M_{i_1,t}$ and $V_{i_2,t} \leq M_{i_2,t}$

Therefore, up to proportionality, the full conditional distribution of $M_{i_1,t}$ takes the form

$$p(M_{i_1,t} = x | \pi_t, n_{i_1,t}, n_{i_2,t}) \propto \binom{n_{i_1,t}}{x} \pi_t^x (1-\pi_t)^{n_{i_1,t}-x} \binom{n_{i_2,t}}{m_{i_1,i_2}-x} \pi_t^{m_{i_1,i_2}-x} (1-\pi_t)^{n_{i_2,t}-m_{i_1,i_2}+x}$$

for $x = x_l, \ldots, x_u$, which defines a discrete distribution on a finite range which may be sampled directly.

4.   Robustness Issues: Analysis of simulated data:

Before addressing the real data, we used a simplified simulation to explore the behaviour of the model in order to assess the impact of different proportions of missing (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 ($\mu_A = 1.49, \sigma_A^2 = 0.35$). 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_A and σ_A^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, 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.
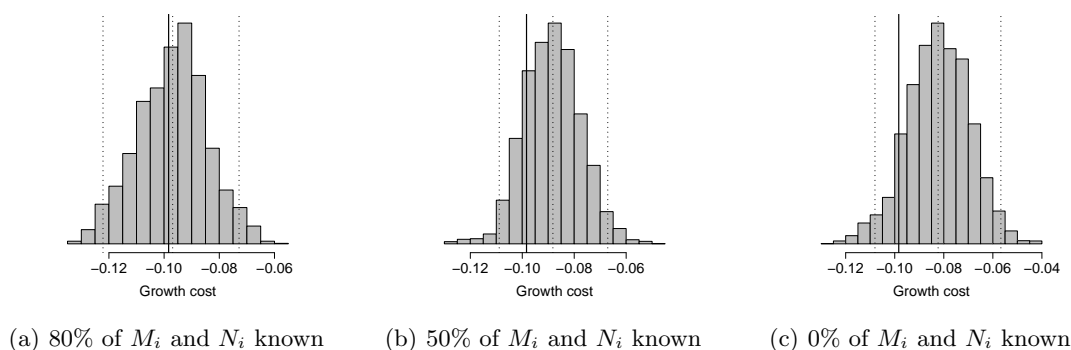
Since inference on missing values denoting vegetative reproduction was a particular focus of this study, we tested our model assuming that different proportions of the N_k and M_k observations were completely known. In three studies, the proportion of individual M_k and N_k values assumed to be known was fixed at 80%, 50% and 0 % respectively; the remaining M_k and N_k assumed to be missing were simulated from using the approaches outlined in section 2.7.2. 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_k and N_k as known.

Results from the simulations (see Figure 3) 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, and that the procedure has good frequentist coverage, that is, the true value often falls within the 95 % posterior credible interval even if no information is available on the N_k and M_k beyond the constraints available in the data. Similar results are obtained for the δ parameter. This result is encouraging, and further confirms the potential of a full Bayesian approach.

5. Analysis of the Wavyleaf Data

Preliminary exploration suggested that site-specific intercepts should be fitted to the growth, flower heads, and size dependent hazard functions; we wished to test whether the extension to the full site-specific model is supported by the observed data. For the wavyleaf data, we therefore considered two types of model:

I *The Across-Site Model* : In this model, intercepts in the various linear predictors are assumed to **vary** across sites, but the coefficients of the continuous covariates are **common** across sites. For example, in the growth model from Section 2.3, we assume a site-specific intercept γ_0, but a common slope γ_1, in the linear predictor. A similar assumption is made for the proportional odds hazards model, with site-specific parameters β_{h_1} and β_{g_1}, and for the number of flowering heads via site-specific η_0 parameters.

(a) 80% of M_i and N_i known   (b) 50% of M_i and N_i known   (c) 0% of M_i and N_i known

Fig. 3. 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 burn-in 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.

In total, ignoring the missing data, this model has a total of 38 parameters when all seven sites are analyzed simultaneously.

II *The Site-Specific Model* : In this model, both intercepts and slopes in the various linear predictors are assumed to **vary** across sites. This model allows us to inspect whether there are any site-to-site differences.

Ignoring the missing data, this model has a total of 62 parameters.

On the basis of ecological prior knowledge, the slopes of number of flowering heads, η_1 in equation (3), were constrained to be non-negative. Note also that the cost on growth and the cost on flowering heads due to ramet production, c_1 and c_2 were presumed to be common parameters across all sites, as were the growth variance parameter σ^2, the culling parameter δ, and the seedling establishment rate λ_{Se}, although these assumptions could also be relaxed.

5.1. Posterior Inference

We implemented the MCMC algorithm described above for the real data set, collecting samples over runs of 50000 iterations after a burn in of 10000 iterations. For the posterior summaries, we thinned the output by taking every twentieth sample; see Appendix A.3 for further details of strategies and convergence assessment. Figure 6 shows growth parameters γ for Model I and Model II for prior values of $N_k \sim Poisson(1)$. Cost parameters for Model I and Model II for prior distributions of $N_k \sim Poisson(1)$, $N_k \sim Poisson(2)$ and $N_k \sim Poisson(5)$ are shown in Figure 4; likewise for δ in Figure 5. Table 2 provides posterior means and standard deviations for all parameters for all models at the second site. Overall posteriors are well defined. The cost parameters and density dependent parameters are sensitive to priors on N_k, but similar in the Across-Sites and Site-Specific Models. Other parameters show no major changes across the models (e.g. Table 2).

Across all models considered, there is evidence that production of a ramet reduces ramet growth in size (the 95 % posterior credible interval for c_1 is entirely in the negative domain), but increases number of flowering heads (the 95 % posterior credible interval for c_2 is entirely in the positive domain, Figure 4). Overall, the models associated with posterior means accurately

(a) All models, c_1 (b) All models, c_2
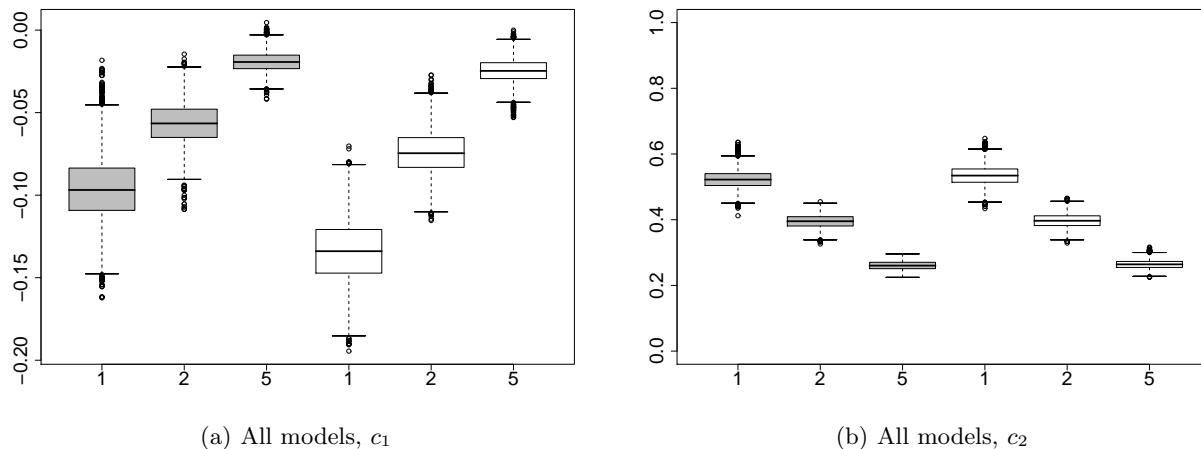
Fig. 4. Posterior boxplots: (a) c_1 in Across-Site (grey, left) and Site-Specific (white, right) models for prior mean values set to $N_k \sim Poisson(1)$, $N_k \sim Poisson(2)$, and $N_k \sim Poisson(5)$ (shown on the x axis); (b) c_2, likewise.

reflect the data (Figure 7). The only slight discrepancy is in the relationship between log size and baseline flowering hazard (Figure 7(c), dashed line), which does not appear to fit well for larger sizes. The reason for this discrepancy appears to be the plateau in the flowering hazard for log sizes above 2.75, which is not captured in the via the logistic link proportional odds model we have suggested; the posterior mean curve shown is essentially determined by two parameters, which are being tied down by the data at the lower log sizes, whereas an extra parameter is needed.

Across all models considered, there is also evidence for strong density dependence acting during ramet establishment in these populations. At the most dense populations observed ($R_t \sim 250$, see Figure 1), even at the high end (e.g. Model II, with the prior set to $N_k \sim Poisson(1)$, see Table 2, Figure 5) only 22 % of new vegetative ramets will successfully establish.

5.2. Model Selection

To compare the (nested) Models I and II formally in this context is problematic due to the presence of the missing data in the formulation. Neither the Bayes Information Criterion (BIC) nor the Deviance Information Criterion (DIC) is ideally suited to model selection when missing data are present (see Celeux et al. 2006). The BIC for standard (non-hierarchical) models equal is equal to

$$-2\log\widehat{\ell} + p\log n$$

where $\widehat{\ell}$ is the maximized likelihood value, p is the number of fitted parameters and $n = 6981$ is the sample size. In our models, it is not clear whether the terms involving the missing data should be included in the likelihood, although as the amount of missing data for the two analyses is identical, the BIC values should not be too misleading.

For the two models fitted, using equation (6) to compute $\log \ell$, approximate BIC values can be computed from the MCMC output. If the largest value of $\log \ell$ observed in the MCMC run is used instead of $\log\widehat{\ell}$, then, for the model in our base analysis where $N_k \sim Poisson(1)$,
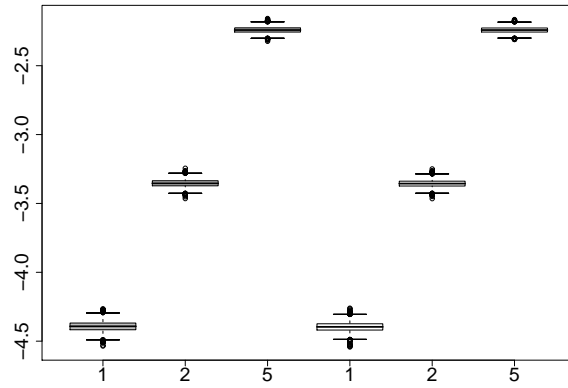
Fig. 5. Posterior boxplots for δ in Across-Site (grey, left) and Site-Specific (white, right) models for prior mean values set to $N_k \sim Poisson(1)$, $N_k \sim Poisson(2)$, and $N_k \sim Poisson(5)$ (shown on the x axis).

(a) Across-Sites Model

(b) Site-Specific Model
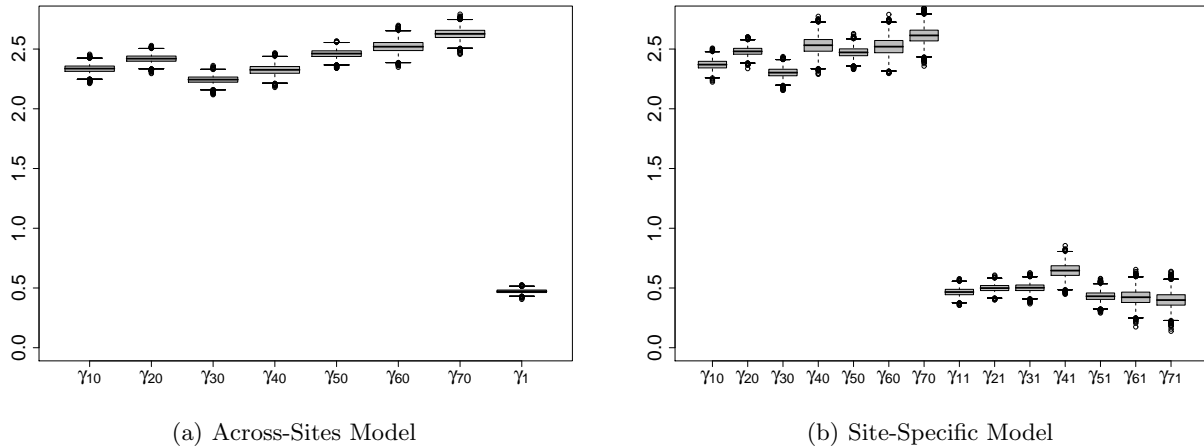
Fig. 6. Posterior boxplots: γ parameters in the (a) Across-Sites and the (b) Site-Specific Models, prior mean is set to $N_k \sim Poisson(1)$

Table 2. Posterior summaries (mean and standard deviation) obtained for the second site using Markov chain Monte Carlo, for Model I and Model II for each of the priors tested, $N_k \sim Poisson(1)$, $N_k \sim Poisson(2)$, and $N_k \sim Poisson(5)$. Underlined parameters are common to all sites in all models.

| | | Model I | | | Model II | | |
| | | Prior | | | Prior | | |
		1	2	5	1	2	5
Density parameter	δ	-4.393	-3.354	-2.241	-4.397	-3.356	-2.241
		(0.037)	(0.027)	(0.022)	(0.036)	(0.027)	(0.021)
Intercept growth	γ_0	2.419	2.433	2.407	2.481	2.489	2.454
		(0.033)	(0.036)	(0.041)	(0.037)	(0.043)	(0.048)
Slope growth	γ_1	0.472	0.475	0.470	0.499	0.504	0.498
		(0.016)	(0.016)	(0.017)	(0.030)	(0.031)	(0.031)
Variance in growth	σ^2	0.193	0.196	0.201	0.183	0.191	0.199
		(0.006)	(0.006)	(0.005)	(0.007)	(0.006)	(0.005)
Ramet cost on growth	c_1	-0.096	-0.057	-0.019	-0.134	-0.074	-0.025
		(0.020)	(0.013)	(0.006)	(0.019)	(0.014)	(0.007)
Intercept head function	η_0	0.826	0.564	0.063	0.822	0.558	0.047
		(0.074)	(0.082)	(0.100)	(0.074)	(0.083)	(0.102)
Slope head function	η_1	1.641	1.650	1.670	1.701	1.691	1.689
		(0.086)	(0.096)	(0.094)	(0.198)	(0.201)	(0.200)
Ramet cost on heads	c_2	0.524	0.395	0.261	0.535	0.397	0.264
		(0.029)	(0.021)	(0.014)	(0.030)	(0.021)	(0.013)
Baseline mort. hazard	h_0	0.208	0.208	0.207	0.193	0.191	0.191
		(0.013)	(0.012)	(0.012)	(0.018)	(0.017)	(0.018)
Baseline flower. hazard	g_0	0.244	0.244	0.247	0.272	0.273	0.271
		(0.023)	(0.022)	(0.024)	(0.027)	(0.029)	(0.026)
Size mortality intercept	β_{h1}	0.070	0.068	0.069	0.148	0.167	0.169
		(0.081)	(0.080)	(0.079)	(0.155)	(0.155)	(0.163)
Size mortality slope	β_{h2}	-0.639	-0.638	-0.641	-0.608	-0.606	-0.606
		(0.049)	(0.049)	(0.050)	(0.099)	(0.101)	(0.100)
Size flowering intercept	β_{g1}	0.225	0.223	0.208	0.075	0.079	0.088
		(0.164)	(0.158)	(0.169)	(0.174)	(0.186)	(0.177)
Size flowering slope	β_{g2}	3.248	3.247	3.253	3.067	3.068	3.077
		(0.126)	(0.127)	(0.128)	(0.252)	(0.258)	(0.253)
Seedling establishment	p_e	0.176	0.176	0.176	0.176	0.176	0.176
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)

(a) Ramet growth data (points) and fitted models from eqn. 1 for $N_k = 0$ (solid line), $N_k = 1$ (dashed line), and $N_k = 3$ (dotted line)
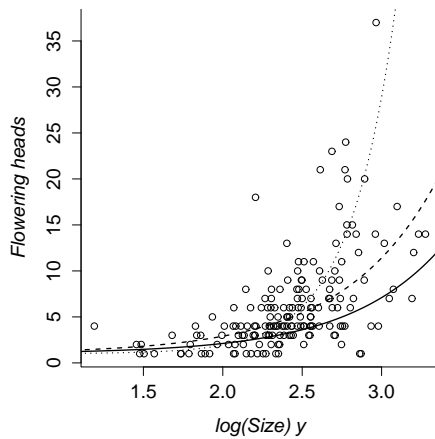
(b) Flowering heads data (points) and fitted models from eqn. 2 for $N_k = 0$ (solid line), $N_k = 1$ (dashed line), and $N_k = 3$ (dotted line)
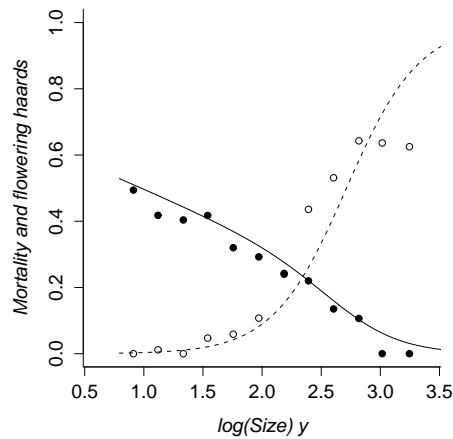
(c) Data for mortality hazards (full points) and flowering hazards (empty points); and fitted models (full and dashed lines respectively)

Fig. 7. Data for Site 2 and predicted fits for models based on posteriors from the Site-Specific model with the prior mean set to $N_k \sim Poisson(1)$ for Site 2 (See Table 2, 4th column).

we have for one typical run

$$\text{Model I} \quad : \quad -2(-21819.32) + 38\log(6981) \approx 43975$$
$$\text{Model II} \quad : \quad -2(-21700.02) + 62\log(6981) \approx 43950$$

These results indicate that Model II provides a globally better fit, but should be treated with caution due to the presence of missing data. For the sensitivity analyses with different prior settings $N_k \sim Poisson(2)$ and $N_k \sim Poisson(5)$, on a typical run we obtained the approximate BIC values displayed below:

Prior Mean	Model I BIC	Model II BIC
1	43974.98	43948.80
2	52422.28	52650.04
5	63183.04	63306.68

These results clearly indicate that the conclusion of the optimal model is prior dependent. The differences arise as the comparisons are effectively being carried out on different data sets, due to the presence of N_k in equations (2) and (3), and the changing priors on these variables, render formal comparison difficult.

The zero-truncated Poisson model from equation (4) was also fitted; from a similarly long run, the BIC for the $N_k \sim Poisson(1)$ prior was estimated to be 44120 (for Model I equivalent) and 43820 (for Model II equivalent), indicating that although is rather less interpretable due to the complicated mean-variance relationship and the relationship between mean and linear predictor, the zero-truncated Poisson model may improve fit for Model II under the $N_k \sim Poisson(1)$ prior. Posterior summaries for the parameters, for the translated and zero-truncated models are included in Table 3. The ecologically substantive inferences remain unchanged.

5.3. Goodness of Fit and Model Adequacy

We also performed a model adequacy assessment for certain model components. A fully Bayesian assessment of model adequacy would be preferable, but although several approaches - based on well-defined posterior expectations, or posterior predictive quantities - have been suggested, there are reservations attached to each of them - see Appendix A.5 for a discussion.
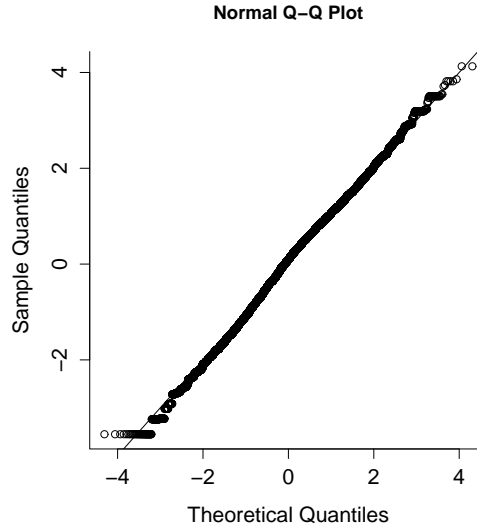
Here we instead adopt the following strategy, and appeal to an asymptotic justification. In large samples, for regular models, any consistent estimator (Bayesian or otherwise) will yield a reasonable estimate of the data-generating parameter, which suggests that rather than computing a posterior expectation by averaging over the posterior samples, we might use a plug-in estimate of the parameter before computing conventional residuals (or p-values). This pragmatic approach is also open to criticism, but yields a readily interpretable result, and reflects more accurately the objectives of the researcher.

Accordingly, to test goodness of fit, we generated a distribution of residuals from the models for growth in size and flowering heads using posterior means for parameters. We mixed over the N_k component by generating a random deviate from the Poisson distribution with mean set by the prior N_k for every observed size (i.e. every individual at every time step), denoted $n_{i,t}$. For growth, residuals for every individual at every time step were generated as

$$(y_{i,t+1} - \widehat{\gamma}_0 - \widehat{\gamma}_1 y_{i,t} - \widehat{c}_1 n_{i,t})/\widehat{\sigma}, \tag{10}$$

Table 3. Posterior summaries (mean and standard deviation) for Model II with the $N_k \sim Poisson(1)$ prior under the translated and zero-truncated models. Note that, due to the different relationships between mean and linear predictor in the two models, direct comparison of parameters η_1, η_2 and c_2 is not straightforward.

		Translated	Zero Truncated
Density parameter	δ	-4.397	-4.397
		(0.036)	(0.036)
Intercept growth	γ_0	2.481	2.494
		(0.037)	(0.057)
Slope growth	γ_1	0.499	0.504
		(0.030)	(0.034)
Variance in growth	σ^2	0.183	0.181
		(0.007)	(0.010)
Ramet cost on growth	c_1	-0.134	-0.140
		(0.019)	(0.032)
Intercept head function	η_0	0.822	1.151
		(0.074)	(0.068)
Slope head function	η_1	1.701	1.439
		(0.198)	(0.171)
Ramet cost on heads	c_2	0.535	0.439
		(0.030)	(0.028)
Baseline mort. hazard	h_0	0.193	0.191
		(0.018)	(0.019)
Baseline flower. hazard	g_0	0.272	0.027
		(0.027)	(0.028)
Size mortality intercept	β_{h1}	0.148	0.155
		(0.155)	(0.161)
Size mortality slope	β_{h2}	-0.608	-0.697
		(0.099)	(0.111)
Size flowering intercept	β_{g1}	0.075	0.057
		(0.174)	(0.181)
Size flowering slope	β_{g2}	3.067	4.097
		(0.252)	(0.373)
Seedling establishment	p_e	0.176	0.175
		(0.007)	(0.007)

Normal Q–Q Plot

Fig. 8. Q-Q plot residuals for Site 2, Site-Specific model for the model for ramet growth in size.

where the $\widehat{\gamma}_0, \widehat{\gamma}_1$, etc. are posterior means. Likewise, for flowering head production, if $\widehat{\lambda}_{S,i,t} = \exp(\widehat{\eta}_0 + \widehat{\eta}_1 y_{i,t} + \widehat{c}_2 n_{i,t})$, and $w_{i,t}$ is the observed number of flowering heads, the residuals were generated as

$$(w_{i,t} - 1 - \widehat{\lambda}_{S,i,t})/\sqrt{\lambda_{S,i,t}} \tag{11}$$

To mix over the prior, we generated five hundred values for $n_{i,t}$ and calculated residuals as above. We then took the resulting full distribution for growth and flowering, and calculated the associated mean and standard deviation. For all sites, for both ramet growth and flowering head production, this did not deviate to any large degree from a $Normal(0,1)$ sample; see Figure 8 for an example.

6. Discussion

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. It also allows explicit parametrization of density dependence of ramet production, despite considerable missing information. This ecological component has key implications for population spread. The posteriors of the parameter δ indicate that density dependence has a considerable effect. For the Site-Specific model with a prior set to $N_k \sim Poisson(1)$ successful establishment of new vegetative ramets has a probability varying between 0.98 when population densities are low and 0.36 when population densities are high. For $N_k \sim Poisson(5)$ these probabilities become 0.82 and 0.03. All the evidence points to population spread through vegetative reproduction decelerating rapidly at high densities, a key piece of inference in understanding ecological systems.

Species' demographic rates are necessarily inter-dependent, 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 growth in size ($c_1 < 0$), the effect on flowering head production is positive ($c_2 > 0$). Unexplained heterogeneity across individuals might explain this pattern. Some individuals might be in spatial areas of high resource availability, allowing them to produce high numbers of both vegetative ramets and flowering heads. In particular, since flowering and ramet production occur at the same time census, instant resource availability per individual could be very important. By contrast, since growth in size is the result of the entire period between census periods, the cumulative effect of resource expenditure is more likely to become apparent, particularly if vegetative ramets remain connected to their parent plants and continue to deplete their parents' resources. This could be explored by including individual effects through a mixed modelling approach. More detailed physiological information on these species would also shed light on this. It would also help clarify which prior specification for N_k should be favoured.

The detection of trade-offs between growth and asexual reproduction (i.e., $c_1 < 0$) has important 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. The sensitivity of forecasting to priors on N_k will both be an important part of this analysis, but also feed back into informing which prior values are the most likely, from the population dynamics patterns obtained.

Statistically, 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.

Both the statistical model used, and the computational strategy seem to be effective in the analysis of the wavyleaf data. The MCMC approach used is especially suited to solving missing data problems where the constraints in the model are relatively complicated. The imputation of the large number of missing values, the numbers of ramets attempted and produced, is perhaps the most problematic step in terms of computation and speed of convergence, and could be improved upon if a simultaneous update of all the missing values could be achieved. In principle, this is straightforward as the full conditional distribution is discrete, but the constraints render the support of this distribution rather awkward, as does its high-dimensionality.

Finally, in our analysis, we have largely restricted ourselves to Poisson models for count data, utilizing log-linear models for the Poisson rates, and with Binomial culling models. In principle, it is straightforward to extend these models to more general ones without complicating the computational approach to any great degree. In other applications and for simulated data, for example, we have implemented models with Negative Binomial likelihoods and plant-specific random effects in the linear predictor. An extension of the constant baseline hazard rate model to a time-dependent baseline hazard is also straightforward to implement, and although this is not useful for the wavyleaf data, it may well be useful in other applications.

Acknowledgements

The authors wish to thank three referees, the associate editor and editor for comments on earlier versions of the paper that have greatly improved the presentation and accessibility of the work. The second author acknowledges the support of an Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant.

Appendices

Appendix A.1.   Notation Summary

- **Data:** for each ramet i in each year t

    - the size $y_{i,t}$ and size the following year $y_{i,t+1}$
    - the survival status $z_{i,t}$ (1=death, 2=flower, 0=neither)
    - the number of flowering heads $w_{i,t}$ if the ramet flowered

- **Data:** for each year t at each site

    - the total number of recruits from seed b_t
    - the total number of recruits from vegetative ramets $\widetilde{m}_t$

- **Missing Data:**   for each ramet i in each year t

    - the number of ramets attempted $N_{i,t}$
    - the number of ramets produced $M_{i,t}$

- **The constraints:**

    - the total number of ramets produced at a site in any year is observed, that is, we observe

$$\sum_{i=1}^{R_t} M_{i,t} = \widetilde{m}_t$$

    where R_t 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 and t, we observe $M_{i,t} > 0$, or equivalently $V_{i,t} = 1$.

Appendix A.2.   The model and likelihood components

In the following we consider data for a single site across all plants i and years t, and let $\mathcal{K} \equiv \{$All combinations of i and $t\}$.

- **Growth in size:** Under the Gaussian autoregressive model, we have

$$Y_{i,t+1}|y_{i,t}, \gamma, \sigma, n_{i,t} \sim Normal\left(\gamma_0 + \gamma_1 y_{i,t} + c_1 n_{i,t}, \sigma^2\right)$$

    Therefore $\ell_1\left(\boldsymbol{y}; \gamma, \sigma, c_1, \boldsymbol{n}\right)$ is given by

$$\ell_1\left(\boldsymbol{y}; \gamma, \sigma, c_1, \boldsymbol{n}\right) = \prod_{t=1}^{11}\left\{\prod_{i=1}^{A_t} f_1\left(y_{i,t+1}; y_{i,t}, \gamma, c_1, \sigma, \boldsymbol{n}\right)\right\}$$

with f_1 the normal density, A_t the total number of individuals that survive in each year t, and $y_{i,t}$ is a measure of ramet size. The initial state of the system, $y_{i,1}$ is taken to be fixed and determined by the observed distribution of sizes at the first time step. In the final year of the study, no data are available for sizes the following year, so $t < 12$.

- **Flowering heads:** For $k \in \mathcal{K}$

$$W_k | y_k, \eta, c_2, n_k \sim 1 + Poisson\left(\exp\{\eta_0 + \eta_1 y_k + c_2 n_k\}\right)$$

As above, $\ell_2\left(\boldsymbol{w}; \boldsymbol{y}, \eta, c_2, \boldsymbol{n}\right)$ is therefore

$$\ell_2\left(\boldsymbol{w}; \boldsymbol{y}, \eta, c_2, \boldsymbol{n}\right) = \prod_{t=1}^{12}\left\{\prod_{i=1}^{F_t} f_2\left(w_{i,t} - 1; y_{i,t}, \eta, c_2, n_{i,t}\right)\right\}$$

with f_2 the Poisson mass function, F_t the total number of flowering ramets in each year t, and $y_{i,t}$ is the log size.

- **Ramets Attempted:** For $k \in \mathcal{K}$

$$N_k | y_k, \theta \sim Poisson\left(\theta\right)$$

Then $\ell_3\left(\boldsymbol{n}; \boldsymbol{y}, \theta\right)$ is given by

$$\ell_3\left(\boldsymbol{n}; \boldsymbol{y}, \theta\right) = \prod_{t=1}^{11}\left\{\prod_{i=1}^{R_t} f_3\left(n_{i,t}; y_{i,t}, \theta\right)\right\}$$

with f_3 the Poisson mass function, R_t the total number of ramets in each year t, and $y_{i,t}$ is a the log size. In the final year of the study, no data are available for total number of ramets attempted and successfully establishing the following year, so $t < 12$.

- **Ramets Observed:** For $k \in \mathcal{K}$,

$$M_k | N_k = n_k, \delta \sim Binomial\left(n_k, \pi\left(\delta, R\right)\right)$$

independently, therefore $\ell_4(\boldsymbol{m}; \boldsymbol{n}, \delta)$ is given by

$$\ell_4(\boldsymbol{m}; \boldsymbol{n}, \delta) = \prod_{t=1}^{11}\left\{\prod_{i=1}^{R_t} f_4\left(m_{i,t}; n_{i,t}, \delta\right)\right\}$$

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

- **The lifetime distribution:**   In the constant baseline hazard model, the distribution of $z_{i,t} | y_{i,t}, \beta_h, \beta_g, h_0, g_0$ has baseline hazards h_0, g_0 modified in a proportional odds model by the log size, $y_{i,t}$. The likelihood contribution for the lifetime component is $\ell_5\left(\boldsymbol{z}; \boldsymbol{y}, \beta_h, \beta_g, h_0, g_0\right)$ is therefore given by

$$\ell_5\left(\boldsymbol{z}; \boldsymbol{y}, \beta_h, \beta_g, h_0, g_0\right) = \prod_{t=1}^{12}\left\{\prod_{i=1}^{R_t} f_5\left(z_{i,t}; y_{i,t}, \beta_h, \beta_g, h_0, g_0\right)\right\}$$

where f_5 is the multinomial term appearing in the product in equation (1), with contribution dictated by the indicator $z_{i,t}$ and individual terms described in Section 2.2. Candidates for h_0 and g_0 were proposed in one of three ways:

(i) from a symmetric Dirichlet distribution with parameters $(1, 1, 1)$ corresponding to the prior distribution for (h_0, g_0).

(ii) using an independence proposal from a Dirichlet distribution chosen from pilot runs.

(iii) using a local Metropolis-Hastings proposal on the logistic scale.

The acceptance probability was altered appropriately for each method, including a term for the Jacobian of the transformation in (iii).

- **Seed establishment:** For each site in any year, assuming that each flowering head in the previous year produces a number of seeds that is $Poisson(\lambda_S)$ distributed independently of all other plants, the total number of seeds produced S, and the number of seedlings that establish, B_t, are related in the following hierarchical model by

$$S \sim Poisson(F\lambda_S) \qquad B_t|p_e, S \sim Binomial\,(S, p_e)$$

where F is the number of flowering heads from the previous year, and where $B_t = b_t$ is directly observed in the data. Marginalizing over the unobserved S, we have that

$$B_t \sim Poisson(F\lambda_S p_e)$$

and hence in the simple model used where p_e is presumed constant across sites, and $\lambda_{Se} = \lambda_S p_e$, the likelihood $\ell_6\,(\boldsymbol{b}; \boldsymbol{w}, \lambda_{Se})$ can be written down explicitly. In the final year of the study, no data are available for total number of seedlings that successfully establish the following year. The full conditional posterior for λ_{Se} can be obtained in closed form; we have that

$$\ell_6\,(\boldsymbol{b}; \boldsymbol{w}, \lambda_{Se}) = \prod_l \frac{(F_l\lambda_{Se})^{b_l}}{b_l!} \exp\{-F_l\lambda_{Se}\} \propto \lambda_{Se}^{\sum_l b_l} \exp\left\{-\sum_l F_l\lambda_{Se}\right\}$$

where l indexes the site/year combinations. Hence the full conditional takes the form

$$\lambda_{Se}|\boldsymbol{b}, \boldsymbol{w} \sim Gamma\left(\sum_l b_l + \alpha_{Se}, \sum_l F_l + \beta_{Se}\right)$$

where $\alpha_{Se} = \beta_{Se} = 1$ are our chosen hyperparameters.

Appendix A.3. MCMC Implementation Details

The MCMC algorithm was implemented according to a tailored version of the strategy described in the main paper (Section 3). The strategy was as follows; we performed pilot runs of the algorithm using only local move Metropolis-Hastings steps, and then restarted the algorithm from high posterior density region and used joint updates of the parameters in blocks, with proposals tuned to capture the posterior correlation amongst the parameters. In the final run, a burn-in of 10000 was used, and then the posterior samples were collected every 20 iterations to further minimize the serial autocorrelation. For most of the parameters, this procedure was sufficient to practically remove all serial correlation from the collected samples. The entire operation was implemented several times to check that the posterior was being sampled appropriately.

For all analyses, we centered the variable y_t to improve performance of the algorithms. Site-specific means are

Site	1	2	3	4	5	6	7
$\overline{y}$	2.46	2.42	2.31	2.57	2.48	2.53	2.53

In every output sample we examined, after the pilot runs and tuning had been implemented, the posterior trace plots had stabilized after 50000 iterations. Formal convergence was assessed using Raftery and Lewis's convergence diagnostic (Raftery and Lewis (1992)) and Heidelberger and Welch's convergence diagnostic (Heidelberger and Welch (1983)), both implemented using the `coda` package in `R` for the MCMC for all parameters. The former is a single chain approach focussed on the precision of estimation of certain posterior quantiles; we used the default `coda` settings. The latter approach consists of testing the null hypothesis that the sampled values come from a stationary distribution, using the Cramer-von-Mises statistic; and then using confidence intervals on the proportion of the chain that passed the stationarity test to verify that the chain is sufficiently long to capture variability in the mean. All tests were passed for all parameters.

Appendix A.4.   *Sampling the Unobserved Ramets Attempted from Section 3.1*

Two methods were used to sample the N_i quantities from their full conditional distribution:

- *Method I:* For each $N_{i,t}$, bearing in mind equation (8), the full conditional mass function for $N_{i,t}$ given $M_{i,t} = m_{i,t}$ is proportional to

$$\ell_1(n)\ell_2(n) \times \frac{\lambda_{i,t}^n \exp\{-\lambda_{i,t}\}}{n!} \times (1 - \pi_t)^n \times \frac{n!}{(n - m_{i,t})!} \tag{12}$$

which is proportional to $\ell_1(n)\ell_2(n) \times g_N(n)$ where

$$g_N(n) = \frac{\{\lambda_{i,t}(1 - \pi_t)\}^n}{(n - m_{i,t})!} \propto \frac{\{\lambda_{i,t}(1 - \pi_t)\}^{n - m_{i,t}}}{(n - m_{i,t})!} \qquad n \geq m_{i,t}.$$

The function g_N defines a probability distribution for $N_{i,t}$ on $\{m_{i,t}, m_{i,t} + 1, ...\}$ that can be sampled directly, as

$$N_{i,t}^{new} \stackrel{\mathfrak{L}}{=} Poisson\left(\lambda_{i,t}(1 - \pi_t)\right) + m_{i,t}$$

Sampling from this model as a proposal for $N_{i,t}$ leaves the acceptance probability as

$$\alpha(N_{i,t}^{old}, N_{i,t}^{new}) = \min\left\{1, \frac{\ell_1\left(N_{i,t}^{new}\right)\ell_2\left(N_{i,t}^{new}\right)}{\ell_1\left(N_{i,t}^{old}\right)\ell_2\left(N_{i,t}^{old}\right)}\right\}.$$

Note that, by construction of g_N, all other terms that are normally present in numerator and denominator of the Hastings ratio cancel: generically, if density function $\pi(x)$ factorizes

$$\pi(x) \propto f(x)g(x)$$

then a Metropolis-Hastings proposal with $q(u, v) \equiv g(v)$ has Hastings ratio

$$\alpha(u, v) = \min\left\{1, \frac{\pi(v)q(v, u)}{\pi(u)q(u, v)}\right\} = \min\left\{1, \frac{f(v)g(v)g(u)}{f(u)g(u)g(v)}\right\} = \min\left\{1, \frac{f(v)}{f(u)}\right\}$$

In our application $f(N) = \ell_1(N)\ell_2(N)$ and $g(N) = g_N(N)$, so the acceptance probability is

$$\min\left\{1, \frac{\ell_1\left(N_i^{new}\right)\ell_2\left(N_i^{new}\right)}{\ell_1\left(N_i^{old}\right)\ell_2\left(N_i^{old}\right)}\right\}.$$

- *Method II:* The full conditional posterior distribution for $N_{i,t}, i = 1, \ldots, R_t$ obtained from (12) is a discrete distribution on $\{m_{i,t}, m_{i,t}+1, \ldots\}$. If we truncate the support at some suitable value (say 20 or 30), then the distribution can be sampled directly using cdf inversion; we can compute the distribution pointwise up to proportionality, and then compute the normalizing constant by summing over the finite range.

Other proposal Metropolis-Hastings mechanisms can be used. Independent proposals subject to the constraints can be readily implemented, with the concomitant change in the acceptance probability. Typically, we would believe that the number of ramets attempted $N_{i,t}$ is not considerably larger than the number of ramets successful $M_{i,t}$, and a Geometric proposal restricted to $N_{i,t} \geq m_{i,t}$ would be suitable. Note that, given that $M_{i,t} = m_{i,t}$, the set of $N_{i,t}$ values are conditionally independent, so can be updated in parallel.

Appendix A.5.   Assessing Model Adequacy using Bayesian p-values

There is no universally accepted method for model assessment in a Bayesian setting; so-called Bayesian p-values (see, for example, Gelman *et al.* (1996), de la Horra and Rodriguez-Bernal (2001)) can be useful exploratory tools, but they can be difficult to interpret. As most typically calculated, the Bayesian p-value is a posterior (predictive) average quantity, and is model and sample size dependent. Generically, for observed data $\boldsymbol{y}$ and parameter θ, the posterior predictive (or Bayesian) p-value is defined as the posterior probability

$$P[D(\boldsymbol{Y}^{rep}, \theta, \boldsymbol{y}) > 0|\boldsymbol{y}] = \int D(\boldsymbol{Y}^{rep}, \theta, \boldsymbol{y})p(\theta|\boldsymbol{y})\, d\theta$$

where $\boldsymbol{Y}^{rep}$ is a hypothetical future data set, comparable with observed data $\boldsymbol{y}$, D is an antisymmetric discrepancy measure. This is a well-defined posterior predictive probability which, intuitively, would be small if the model (likelihood and prior combined) does not predict the data well. This quantity is readily computed from MCMC output and is a useful exploratory tool. In any specific application, the user is required so specify the discrepancy function D, and also potentially the portion of $\boldsymbol{y}$ that is to be used for comparison. For example, for the wavyleaf data, we might consider the log size data component, and use a discrepancy based on the sum of prediction errors across the entire data set

$$D(\boldsymbol{y}^{rep}, \theta, \boldsymbol{y}) = \sum_{i,t}\left(y_{i,t}^{rep} - y_{i,t}\right) \tag{13}$$

where the collection of $y_{i,t}^{rep}$ values are sampled at each MCMC iteration from the conditional distribution given as before, in the autoregressive model, as

$$Y_{i,t}^{rep}|y_{i,t-1}, \gamma, \sigma, n_{i,t-1} \sim N\left(\gamma_0 + \gamma_1 y_{i,t-1} + c_1 n_{i,t-1}, \sigma^2\right)$$

By construction, the distribution of the random quantity $D(\boldsymbol{Y}^{rep}, \theta, \boldsymbol{y})$ should be centered at zero, and the Bayesian p-value should be near zero or one only if there is an appreciable

difference between $\boldsymbol{y}^{rep}$ and $\boldsymbol{y}$, that is, if the observed data and the hypothetical data generated under the model differ greatly. It is usually straightforward to propose and compute such a suitable discrepancy for different aspects of the observed data.

We computed the Bayesian p-value for the the wavyleaf data, under Model II with the $Poisson(1)$ prior for N_k. We implemented this approach with the discrepancy in equation (13) for the log size data, and obtained a Bayesian p-value of 0.50 with a Monte Carlo standard error of 0.01. The distribution of $D(\boldsymbol{Y}^{rep}, \theta, \boldsymbol{y})$ is displayed in Figure 9.
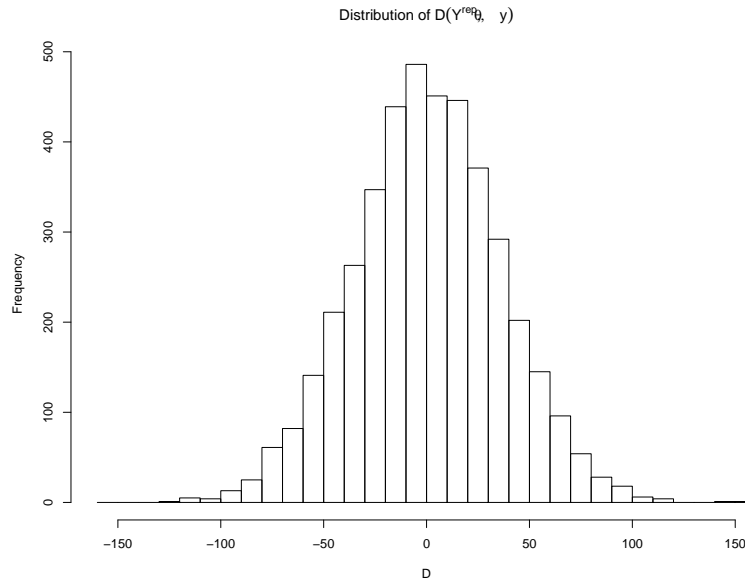
Fig. 9. Distribution of sampled $D(\boldsymbol{Y}^{rep}, \theta, \boldsymbol{y})$

However, despite being suitable in certain circumstances for identifying major model mis-specification (see Gelman (2003) for persuasive support), Bayesian p-values remain contro-versial. In addition to the potential difficulties of implementation (for example, the choice of discrepancy), the mathematical properties of the Bayesian p-value are somewhat problematic; see Hjort *et al.* (2006) and Bayarri and Castellanos (2007) for a discussion and extensions. Unlike the frequentist p-value, which is uniformly distributed if the fitted model is correct (and under standard asymptotic assumptions), the Bayesian p-value has, in general, no specific dis-tribution even if the model being fitted is the correct (data-generating) model. For small data sets, bootstrap methods can be used to calibrate the p-value correctly, as its distribution un-der hypothetical replicate data sets can be constructed, but this is extremely computationally demanding and not feasible for large data sets. Overall the Bayesian p-value is not well-suited to formal model assessment.

Similar arguments apply to "Bayesian residuals"; a possible (and commonly implemented) MCMC strategy is to compute classical residuals at each iteration, and to store and then summarize the resulting posterior samples of the residuals; again, the distribution of the resulting samples, or sample summaries, is not straightforward to understand.

In both cases, it is the Bayesian posterior averaging that is causing difficulties in interpre-tation. Thus, despite being potentially useful as an informal diagnostic, there is little formal (decision-theoretic) justification for the use of Bayesian p-values.

REFERENCES

Bayarri M. J. and Castellanos, M. E. (2007) Bayesian checking of the second levels of hierarchical models, *Statist. Sci.*, 22, 3, 363-367.

Brooks, S. P., Catchpole, E. A., Coulson, T., Lebreton, J.-D. & Morgan, B. J. T. (2004) Recent progress in using marked individuals to understand animal population biology. *Technical Report: 2004/UKC/IMS/04/26*

Brooks, S. P., Catchpole, E. A., Morgan, B. J. T. & Harris, M. P. (2002) Bayesian methods for analysing ringing data. *J. Appl. Statist.*, **29**, 187 - 206.

Cam, E., Link, W. A., Cooch, E. G., Monnat, J.-Y. & Danchin, E. (2002) Individual covariation in life history traits: seeing the trees despite the forest. *American Naturalist*, **159**, 96 - 105.

Celeux, G., Forbes, F., Robert, C.P. and Titterington, D. M. (2006) Deviance information criteria for missing data models. *Bayesian Analysis*, **1**(4), 651–674

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., Oguge, 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. *American Naturalist*, **167**, 410 - 428.

Gardner, S. N. & Mangel, M. (1999) Modeling investments in seeds, clonal offspring, and translocation in a clonal plant. *Ecology*, **80**, 1202-1220.

Gelman, A. (2003) A Bayesian Formulation of exploratory data analysis and goodness-of-fit testing. *Internat. Statist. Rev.*, 71, 2, $369 - 382$.

Gelman, A., Meng, X.-L. & Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statist. Sinica* 6 733807.

Heidelberger, P. and Welch P. D. (1983) Simulation run length control in the presence of an initial transient. *Ops. Res.*, 31, 1109-44

Hjort, N. L., Dahl, F. A. and Steinbakk, G. H. (2006), Post-processing posterior predictive p-values, *J. Amer. Statist. Assoc.*, 101, 1157-1174.

de la Horra, J and Rodriguez-Bernal, M. T. (2001) Posterior predictive p-values: what they are and what they are not, *Test*, 10, 1, 1133-0686.

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. *American Naturalist*, **168**, 76 - 87

Morgan, B. J. T. (2000) *Stochastic Modelling*, Arnold, London.

METCALF, J.C., Rose, K.E. & Rees, M. (2005) The evolutionary demography of monocarpic perennials. *Trends Ecol. Evol.*, **18**, 471 - 480

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.

RAFTERY, A. E. and Lewis, S. M. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statist. Sci.*, 7, 493-497.

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**, 453465.