·17·

Models for demography of plant populations

James S. Clark, Dave Bell, Michael Dietze, Michelle Hersh, Ines Ibanez, Shannon L. LaDeau, Sean McMahon, Jessica Metcalf, Emily Moran, Luke Pangle and Mike Wolosin

17.1 Introduction

Ecologists seek to understand how demographic rates contribute to species diversity. Birth, growth, and survival together determine population growth. Demographic rates are related to one another, and they depend in complex ways on environmental variables. At a given age, an organism allocates energy in ways that affects current and future growth, fecundity, and survival risk. These relationships change through time as resource availability changes, and organisms develop and age. For the population ecologist challenges include inference not only on specific demographic rates, but also on how they combine to determine population growth. In this chapter we discuss how hierarchical Bayes analysis can help synthesize information from a range of sources to understand how demographic rates relate to one another and might contribute to biodiversity.

The key challenges to demographic inference in ecology include the availability of many sources of incomplete information, often measured at different scales, and the large number of interactions among demographic components. Population growth depends on all demographic rates, birth, growth, death, and migration. Each of these rates can respond to a fluctuating environment, including other organisms. Many factors are observable or only weakly related to factors that can be measured. For example, rarely can the cause of death be determined for organisms in the wild. Without knowledge of cause it is difficult to isolate and model individual risk factors, let alone how they interact. Even where cause might be identified, challenges remain. If death could be attributed to, say, 'drought' (Condit *et al.* 1995, Suarez *et al.* 2004, Nepstad *et al.* 2007, van Mantgem and Stephenson 2007), a series of questions arise: is it the daily, seasonal, or annual average soil moisture that is most important? Is it duration or intensity? Which interactions determine why only a fraction of the population died? We rarely have the information to address such questions, and we often lack understanding of the important scales (e.g. weekly or monthly drought?). These challenges have necessitated a superficial view

of demography that is focused on annual rates with limited connection to covariates.

To motivate some of the complexity that follows, consider limitations of current demographic models. Traditionally, most demographic analyses include no predictors beyond age or size (i.e. the standard age-structured model (Leslie matrix) or stage-structured model (see the reviews of Caswell 2001, Gurevitch *et al.* 2002)). Models involving covariates tend to include a demographic rate as a response variable (offspring born per female per year, annual growth rate, fraction of the population making a transition to a different stage of life, survival probability) and, perhaps, a small number of predictors. The conclusions that can be drawn from such models are limited, because it is recognized that they may not accommodate important factors affecting the data, even for simple experiments. For example, seed production of trees is rarely directly observed in natural populations, because it cannot be quantified in forest canopies. The covariates are often represented by crude indices. The standard assumption that error should enter as a stochastic envelope around a deterministic function of predictor variables, such as

$$y_i = x_i'\beta + \varepsilon_i$$

implies that x_i is known much better than y_i. If x_i is a GIS layer, an instrument that records with error, an incubation culture of biological activity, a classification scheme based on unreliable detection, or the interpretation of a fuzzy image, we might better represent the problem by including a model for x. If so, we need to consider how to coherently connect a model for x and a model for y and how to allow for uncertainty and still estimate everything. Growth rates are usually measured with less error than are the environmental covariates used as predictors, including such difficult-to-quantify resources as solar radiation reaching the partially shaded crown, soil moisture, and nutrient supply, having substantial spatiotemporal variation that is never well quantified (Beckage and Clark 2003, Kobe *et al.* 2006, Mohan *et al.* 2007). If mortality is modelled as a function of growth, and observations are not available until after death, the growth rates are rarely available, requiring either annual measurements on trees that both survive and die up until the time of death or increment cores of trees (which are too laborious to obtain on large numbers of individuals). Either way, a model is needed to account for the way in which the covariate data were collected (Kobe *et al.* 1995, Wyckoff and Clark 2000). Moreover, we are now considering multiple demographic rates (growth and survival), connected by virtue of the fact that individuals near death may allocate less to growth.

As ecologists increasingly want comprehensive inference that derives from multiple underlying processes connecting inputs and responses, both of which are partially known, the challenge comes in devising ways to synthesize the

sources of information, allow for observation error and uncertainty in the underlying model itself. How can we capture interactions between individual models for growth, fecundity, and mortality risk? Doing so requires that they be fitted simultaneously, as part of an integrated model of demographic change. The model must reflect the uncertainties that enter through both process and data at multiple stages. Useful inference requires that we exploit information coming not only from data, but also from theory and previous observations and experiments.

Multistage models provide a natural framework for organizing how ecologists think about inference. Bayesian techniques provide a natural approach for analysis of such models. Implementation requires that they can be structured to provide transparency regarding assumptions, model behaviour, and parameter estimates. For example, does the process model capture the relationships in realistic ways, and would we recognize failure to do so? Does the model make realistic predictions at all levels, including for state variables and observations at different scales? The application presented here highlights ways to integrate information for demographic inference, connecting models for important components of the problem, each allowing for uncertainty. We model fecundity, dispersal, growth, and survival with covariates that include some of these demographic rates and light availability, a key resource that limits plant growth. We show how the large number of estimates that come from the analysis can be summarized synthetically to provide deeper insight about relationships among demographic rates and how they respond to covariates.

17.2 Demographic data

We use data collected from tree plots $j = 1, \ldots, J$ and covariates to infer demographic rates. Data include observational studies and whole-stand manipulations, which allow us to break up correlations in some of the important covariates. Study areas include $J = 9$ plots of mapped forest stands in the Piedmont (Duke Forest) and southern Appalachian Mountains (Coweeta) of North Carolina, USA. The plots were selected to span a range of topographic, soil moisture, soil types, and elevation characteristics (Table 17.1), supplemented with experimental manipulations. In this study we report observational data come from trees $i = 1, \ldots, I_j$ on plots j in years $t \subset \{t_{ij}, t_{ij} + 1, \ldots, T_{ij}\}$. Plots were established in years $t_j \subset \{1991, \ldots, 2000\}$, when trees were first mapped, identified to species, and measured for diameter. The first observation year for individual ij is the year when plot j was established, t_j, or when the tree grew to sufficient size (2 m in height) to be measured, whichever came first. The last observation year $T_{ij} \leq 2008$ is the last year the individual was observed, at which point it might still be alive or not. Species codes are listed

Table 17.1 Plot characteristics and and number of trees by plot. Species codes are listed in Table 17.3.

Plot	C118	C218	C318	C427	C527	CLG	CUG	DBW	DHW	Total
Elev (m)	780	820	870	1110	410	030	140	70	70	
Lat/Long	35°03′N. 83°27′W							35°58′36″N 79°5′48″W		
Soil type	Typic & humic hapludults, typic dystrochrepts, typic haplumbrepts							Typic & oxyaquic vertic hapludalfs		
First year	1992	1992	1992	1992	1992	2000	2000	2000	1999	
Area (ha)	0.64	0.64	0.64	0.64	0 64	2.75	1.45	4.11	2.40	13.91
acru	496	136	211	257	14	982	608	2761	666	6131
acsa	0	2	1	0	82	35	0	13	0	133
acpe	5	223	15	15	230	451	21	0	0	960
acba	0	0	0	0	0	0	0	4	114	118
acun	0	0	0	0	0	13	0	2	0	15
beal	0	1	2	0	157	0	0	0	0	160
bele	5	38	12	4	66	8	1	0	0	134
beun	0	0	0	0	0	29	38	0	0	67
caca	0	0	0	0	0	0	0	146	439	585
cagl	56	33	41	15	5	36	10	153	31	380
caov	0	0	2	0	0	0	0	0	66	68
cato	1	0	0	0	0	48	3	395	51	498
caun	1	0	1	2	0	162	5	77	73	321
ceca	0	0	0	0	0	0	0	282	39	321
cofl	45	78	27	10	3	118	14	1405	501	2201
fram	0	0	1	4	63	46	0	658	802	1574
list	0	0	0	0	0	0	0	1523	648	2171
litu	11	70	12	9	0	654	6	371	161	1294
nysy	103	30	113	117	0	282	335	361	457	1798
piri	36	0	0	0	0	0	0	0	0	36
pist	2	16	0	0	0	0	0	0	0	18
pita	0	0	0	0	0	0	0	391	175	566
piec	0	0	0	0	0	0	0	76	1	77
pivi	0	0	0	0	0	0	0	25	1	26
qual	12	0	0	0	1	0	0	231	65	309
quco	25	7	1	14	0	51	44	0	0	142
qufa	0	0	0	0	0	0	0	21	5	26
quma	10	0	0	0	0	0	0	20	0	30
quph	0	0	0	0	0	0	0	18	73	91
qupr	91	34	102	101	0	180	173	0	0	681
quru	38	5	21	57	24	139	49	65	0	398
qust	0	0	0	0	0	0	0	41	41	82
quve	45	6	7	3	0	12	14	30	0	117
quun	0	0	0	0	0	18	18	4	0	40
rops	24	6	4	12	0	95	30	1	0	172
tiam	0	5	0	0	87	0	0	0	0	92
tsca	8	27	8	74	3	243	38	0	0	401
ulal	0	0	0	0	0	0	0	719	558	1277
ulam	0	0	0	0	0	0	0	69	147	216
ulru	0	0	0	0	0	0	0	0	64	64
ulun	0	0	0	0	0	0	0	139	86	225
Total trees	1014	717	583	694	735	3602	1407	10001	5264	24015
Seed traps	20	20	20	20	20	73	43	128	66	410
Total seeds	6413	46964	22685	13298	334374	20814	10801	75481	23487	554317

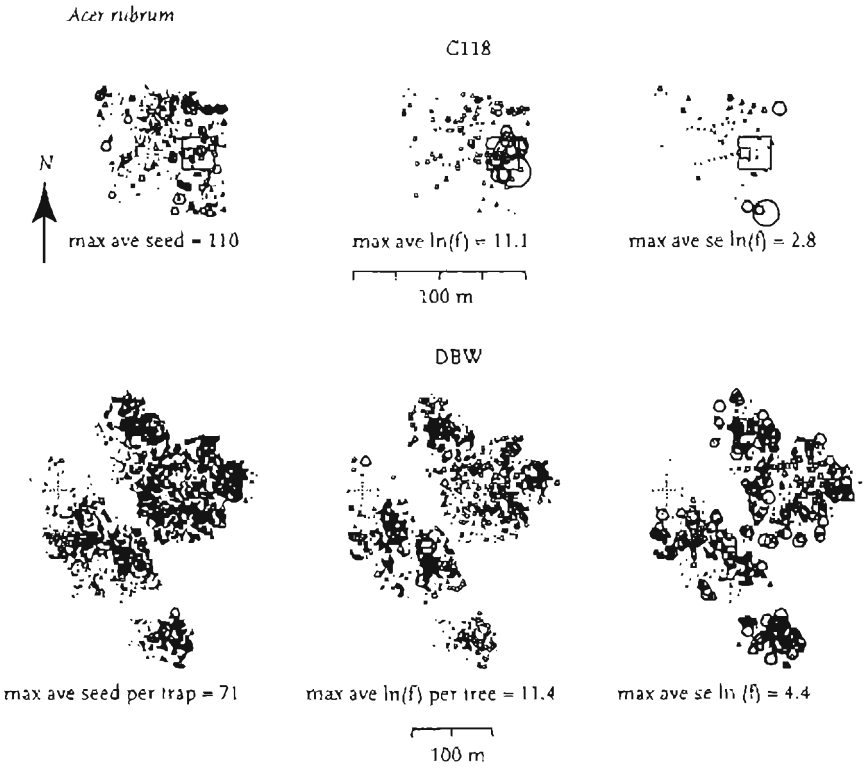

Fig. 17.1 Examples of two mapped stands used for demographic inference, showing stems of a single genus, *Acer*, represented by circles. Boxes are shown at seed trap locations with box size proportional to average annual seed collection for the entire study period. Plot C118 is small and has not been manipulated. Plot DBW is larger (note scale bars), with seed traps clustered in and around the locations of eight canopy gaps created in 2002. From left the series of three maps for both locations have circle size scaled to (1) stem diameter, (2) mean estimate of tree fecundity, and (3) standard error of tree fecundity estimates (not to be confused with the standard deviation in fecundity over time).

in Table 17.2. Example maps of two study plots are shown in Figure 17.1. Environmental data include static variables (elevation, slope, aspect) and variables that fluctuate over time, with additional variation among plots (temperature) and within plots (soil moisture, light penetration). This analysis focuses on how light availability and tree size affects demographic rates, including relationships among them.

A subset of the mapped plots was used for canopy gap experiments. Following collection of several years of pretreatment data, trees were removed from the canopy in 20 m or 40 m wide patches to simulate the small and large treefall gaps that occur when a single tree falls or when groups of trees fall, as during storms. Canopy trees were pulled down using a skidder or bulldozer located outside the plot and left in position, some having snapped, others uprooted.

Table 17.2 Species codes used in other tables and figures.

Code	Species	Code	Species
acru	*Acer rubrum*	piri	*Pinus rigida*
acsa	*Acer saccharum*	pist	*Pinus strobus*
acpe	*Acer pensylvanicum*	pita	*Pinus taeda*
acba	*Acer barbatum*	piec	*Pinus echinata*
acun	*Acer unknown*	pivi	*Pinus virginiana*
beal	*Betula alleghaniensis*	qual	*Quercus alba*
bele	*Betula lenta*	quco	*Quercus coccinea*
beun	*Betula unknown*	qufa	*Quercus falcata*
caca	*Carpinus caroliniana*	quma	*Quercus marilandica*
cagl	*Carya glabra*	quph	*Quercus phellos*
caov	*Carya ovata*	qupr	*Quercus montana*
cato	*Carya tomentosa*	quru	*Quercus rubra*
caun	*Carya unknown*	qust	*Quercus stellata*
ceca	*Cercis canadensis*	quve	*Quercus velutina*
cofl	*Cornus florida*	quun	*Quercus unknown*
fram	*Fraxinus americana*	rops	*Robinia pseudoacacia*
ilde	*Ilex decidua*	tiam	*Tilia americana*
ilop	*Ilex opaca*	tsca	*Tsuga canadensis*
list	*Liquidambar styraciflua*	ulal	*Ulmus alata*
litu	*Liriodendron tulipifera*	ulam	*Ulmus americana*
nysy	*Nyssa sylvatica*	ulru	*Ulmus rubra*
oxar	*Oxydendron arboreum*	ulun	*Ulmus unknown*

Damage to canopy trees and understory trees (snapped or bent by pulled trees) was recorded as the basis for analysis of damage effects on growth (Dietze and Clark 2008). For the analysis presented here, manipulation had the greatest effect on light availability, represented by 'exposed canopy area'.

To allow for inference on demographic interactions, we use a structure that combines important relationships but is constrained by what can be observed or inferred. An individual is characterized by several state variables, some of which are constant, some change over time, some are continuous, and others discrete. State variables differ in terms of how directly each can be observed. Trees are classified according to genus (e.g. *Quercus* for oak) and species (*Q. rubrum* for red oak). The genus to which a tree belongs is known. The species is also taken to be known, in the sense that it will not be inferred, but seeds are confidently identified only to genus. For this reason, trees of the same genus are modeled together. The genera and species are summarized in Table 17.3. There is an unknown species class for trees from several genera, where large individuals could not be confidently ascribed to a particular species (Table 17.3). Because many seeds can only be classified to the level of genus, we infer seed production as the combined contributions from all trees in that genus. For example, the analysis for *Quercus* includes 10 species, because not all acorns could be identified to species level (Clark *et al.* 1998, 2004), whereas the analysis of

Table 17.3 Seed-trap years (same for all taxa) and tree years, grouped by plot and by genus, as analysed here. Species codes are listed in Table 17.3.

Plot	C118	C218	C318	C427	C527	CLG	CUG	DBW	DHW	Total
Trap year	320	320	320	320	320	490	301	1016	342	3749
					Acer					
acru	8432	2312	3587	4369	238	8838	5472	24849	6660	64757
acsa	0	34	17	0	1394	315	0	117	0	1877
acpe	85	3791	255	255	3910	4059	189	0	0	12544
acba	0	0	0	0	0	0	0	36	1140	1176
acun	0	0	0	0	0	117	0	18	0	135
					Betula					
beal	0	17	34	0	2669	0	0	0	0	2720
bele	85	646	204	68	1122	72	9	0	0	2206
beun	0	0	0	0	0	261	342	0	0	603
					Carpinus					
caca	0	0	0	0	0	0	0	1314	4390	5704
					Carya					
cagl	952	561	697	255	85	324	90	1377	310	4651
caov	0	0	34	0	0	0	0	0	660	694
cato	17	0	0	0	0	432	27	3555	510	4541
caun	17	0	17	34	0	1458	45	693	730	2994
					Cercis canadensis					
ceca	0	0	0	0	0	0	0	2538	390	2928
					Cornus florida					
cofl	765	1326	459	170	51	1062	126	12645	5010	21614
					Fraxinus americana					
fram	0	0	17	68	1071	414	0	5922	8020	15512
					Liquidambar styraciflua					
list	0	0	0	0	0	0	0	13707	6480	20187
					Liriodendron tulipifera					
litu	187	1190	204	153	0	5886	54	3339	1610	12623
					Nyssa sylvatica					
nysy	1751	510	1921	1989	0	2538	3015	3249	4570	19543
					Pinus					
piri	612	0	0	0	0	0	0	0	0	612
pist	34	272	0	0	0	0	0	0	0	306
pita	0	0	0	0	0	0	0	3519	1750	5269
piec	0	0	0	0	0	0	0	684	10	694
pivi	0	0	0	0	0	0	0	225	10	235
					Quercus					
qual	204	0	0	0	17	0	0	2079	650	2950
quco	425	119	17	238	0	459	396	0	0	1654
qufa	0	0	0	0	0	0	0	189	50	239
quma	170	0	0	0	0	0	0	180	0	350
quph	0	0	0	0	0	0	0	162	730	892
qupr	1547	578	1734	1717	0	1620	1557	0	0	8753
quru	646	85	357	969	408	1251	441	585	0	4742
qust	0	0	0	0	0	0	0	369	410	779

(cont.)

Table 17.3 (*Continued*).

Plot	C118	C218	C318	C427	C527	CLG	CUG	DBW	DHW	Total
quve	765	102	119	51	0	108	126	270	0	1541
quun	0	0	0	0	0	162	162	36	0	360
					Robinia pseudoacacia					
rops	408	102	68	204	0	855	270	9	0	1916
					Tilia americana					
tiam	0	85	0	0	1479	0	0	0	0	1564
					Tsuga canadensis					
tsca	136	459	136	1258	51	2187	342	0	0	4569
					Ulmus					
ulal	0	0	0	0	0	0	0	6471	5580	12051
ulam	0	0	0	0	0	0	0	621	1470	2091
ulru	0	0	0	0	0	0	0	0	640	640
ulun	0	0	0	0	0	0	0	1251	860	2111

Carpinus includes a single species (Table 17.3). Trees retain the species identity, having some parameters that are species-specific, whereas seeds are modelled as being potentially produced by trees of the entire genus on a probabilistic basis. This approach allows combination of observations at the scale of individual trees and at seed-trap scale (a seed traps accumulate seeds from all trees simultaneously).

In addition to diameter and species, individual level observations include survival, canopy status, and reproductive maturation status. Canopy status involved ordinal classes derived from standard classifications used in forestry. At one to four year intervals, individuals were assigned to:

Class 1: suppressed in the understory, with access limited to sunflecks (e.g. intermittent patches of direct sunlight);
Class 2: intermediate, with not more than 20% of the canopy exposed to some direct sunlight;
Class 3: codominant, with > 20% of the canopy exposed to direct sunlight during part of each day.

A suppressed individual in the understory would be assigned Class 1 status, but could change to Class 3 status if it occupied a canopy gap following loss of the overstory.

Additional information on canopy status comes from low-altitude aerial photo coverage of plots, used to segment and measure canopy areas of trees visible from above. The modelling of canopy exposure based on status and remote sensing observations, combined with allometric models of canopy area is described in Valle *et al.* (2009). Posterior means and variances from that analysis are used as prior means and variances for the analysis presented here.

Maturity status observations were made at irregular years during the flowering season (taxa such as *Ulmus* and *Acer rubrum* have conspicuous flowers before leaf-out in spring), the fruiting season (fruits often visible in the lower canopy include *Cornus florida, Cercis canadensis*), and winter (lack of leaves permits identification reproductive structures of *Liriodendron tulipifera, Liquidambar styraciflua*). Classes of observations are:

uncertain: the largest class, because fruits and/or flowers are difficult to observe;
not mature: if the entire canopy could be clearly observed to have no fruiting structures during the known flowering and/or fruiting season;
flowering: establishes that the individual is mature;
seeds/fruits present: establishes maturity and, for dioecious species, female status;
female or male flowers present: in rare cases individuals of a dioecious species could be assigned gender on the basis of flower structure.

The observations are summarized by maturation status and gender status in Table 17.4, where $\theta_{ij,t} = p(Q_{ij,t} = 1)$ is the probability that individual i on plot j is mature in year t, $v = p(q_{ij,t} = 1 | Q_{ij,t} = 1)$ is the probability that a mature individual will be identified as such, $\phi = p(H_{ij} = 1)$ is the fraction of individuals that are female for the entire population. The probabilities of $Q|q$ and $H|h$ are provided in Table 17.4 primarily for explanatory purposes, because data models discussed in Section 17.3 involve multiple observations. Note that statuses must be modeled only for Table 17.4 entries not containing a zero or a one.

Table 17.4 Indicators and probabilities of maturity and gender conditional on observations.

Observation	Maturity indicator $q_{ij,t}$	Maturity probability[1] $\Pr(Q_{ijt} = 1 \mid q_{ij,t})$	Gender indicator $h_{ij,t}$	Gender probability[1,2] $\Pr(H_{ij} = 1 \mid h_{ij,t})$
no observation	—	$\theta_{ij,t}$	—	ϕ
uncertain	0	$\frac{(1-v)\theta_{ij,t}}{1-v\theta_{ij,t}}$	0	ϕ
not mature	−1	0	0	ϕ
flowering	1	1	0	ϕ
seeds/fruits	1	1	1	1
male or female flowers	1	1	0 or 1	0 or 1

[1] The probabilities are shown for the special case that there is a single observation per individual. In fact, there are multiple status observations, modelling of which is discussed in Section 17.3.1. Symbols are θ – probability of being in the mature state, v – probability of detecting mature status, ϕ – female fraction of the population.
[2] Gender status is assumed static (H has subscript ij), whereas there are multiple observations of status for each individual (h has subscript ij,t).

Unfortunately, most observations are 'uncertain', requiring that most statuses must be modelled.

Seed trap data are the basis for fecundity estimates, and they further contribute to estimates of maturation and gender statuses. Seeds cannot be counted in the dense canopies characteristic of closed forests, but spatiotemporal seed data can be used to model fecundity (Clark *et al.* 1998, 1999, 2004). Fecundity is estimated from seed traps $k = 1, \ldots, K_J$ using models of dispersal (Clark *et al.* 1998, 1999, 2004), where K_J ranged from 20 to 128 seed traps. Seed traps were emptied from two–four times annually, seeds identified to genus or species (some seeds can only be identified to the genus level), and counted. For modelling purposes, seed data were accumulated to total per trap per year. $s_{kj,t}$ (Table 17.3).

Diameter measurements and increment cores provide information on individual tree growth. Because diameter fluctuates with stem moisture content, and diameter measurements have error, we measured diameters $D_{ij,t}$ at 1–4 year intervals. In addition to diameter measurements, increment cores were extracted from a subset of trees, providing a record of past growth $d_{ij,t} = D_{ij,t+1} - D_{ij,t}$ for the individual up to the year in which the core was extracted. Modeling of diameter-census and increment-core data are described by Clark *et al.* (2007). Posterior means and variances for each tree year obtained from that analysis are used as priors for the analysis presented here.

17.3 Models to synthesize data and previous knowledge

Consider a forest containing trees of different species, gender, age, and size, each experiencing the local environment in ways that depend on some factors that can be measured and others that cannot; here we focus on light availability. The responses of interest include growth rate, gender ratio, maturation status, fecundity, seed dispersal, and survival risk. Combinations of these demographic rates within individuals determine the growth rates of populations and, thus, community biodiversity.

An individual's response to the environment produces variation in demographic rates. Resource availability (here we consider light) contributes to overall health. Resources vary at many scales, depending on supply and on competition with neighbours. For example, light levels vary throughout the day and seasonally, and they are reduced by nearby trees that shade one another. Fine scale heterogeneity is most obvious where canopy gaps form, allowing from 10 to 100% of full sunlight to penetrate to the forest floor, depending on gap size. By contrast, the uninterrupted canopy intercepts 95 to 99% of incoming radiation. Due to spatial heterogeneity, individual trees are exposed to conditions that differ from their neighbors, and these differences can change

over time with changes in canopy structure and as interannual climate variation moderates the impact of resource supply. Some of these factors can be measured in the environment, but it is important to also allow for variation that cannot be ascribed to measurable factors. Here we describe the model for demographic rates.

17.3.1 Gender and maturation

The gender of a tree remains constant, whereas maturation status changes from immature to mature over time. Maturation status and gender can be confirmed for some trees (Table 17.4). Presence of seeds indicates maturity and (for dioecious species) female status. The presence of flowers indicates maturity, but it does not mean that an individual belonging to a dioecious species is female, unless it can be identified as a female flower. When the entire crown can be observed, lack of flowers or fruits during the flowering/fruiting season is taken to indicate immaturity. In crowded stands, absence of reproductive effort can rarely be confirmed by such observations, so detection is uncertain.

The transition from immature to mature is treated as a hidden Markov process. Modelling of gender and maturation is complicated by the fact that probabilities depend on the entire history of observations on individual ij. Consider an individual that is observed once. If maturation status is uncertain ($q_{ij,t} = 0$ in Table 17.4), then the individual is mature with probability $(1 - v)\theta_{ij,t}/(1 - v\theta_{ij,t})$. But additional observations complicate the model. For example, mature status is more likely for an individual last known to be immature 10 year ago than it is for an individual last known to be immature one year ago. Likewise, an individual is more likely to be mature in year t if it is first known to be mature in year $t + 1$ than if it is first known to be mature in year $t + 10$. Furthermore, an individual observed to be of unknown status once is more likely to be mature than is an individual observed to be of unknown status 10 times. In other words, modelling of status must accommodate not only the probabilities contained in Table 17.4, but also how they must be combined to accommodate the differing observation histories of each individual. Here we discuss these probabilities. We first discuss conditioning on observations listed in Table 17.4, followed by seed data.

The unconditional probability that individual i on plot j is female is termed the female fraction $\Pr(H_{ij} = 1) = \phi$, (Table 17.2) and the probability of being male $\Pr(H_{ij} = 0) = 1 - \phi$. Observations of gender status were obtained at irregular intervals $h_{ij,t}$. If the individual is observed to be female $h_{ij,t} = 1$ or male $h_{ij,t} = 0$ then

$$\Pr\left(H_{ij} = 1 \middle| h_{ij,t} = 1\right) = 1$$

$$\Pr\left(H_{ij} = 0 \middle| h_{ij,t} = 0\right) = 1.$$

In other words, gender is only assigned if it is certain. If there are no observations for gender, information enters solely though seed rain data. If seed density near an individual is high, then the probability that it is female is large, and vice versa. Thus, fecundity and maturation must be modeled together.

The unconditional probability that the individual is in the mature state $\Pr(Q_{ij,t} = 1) = \theta_{ij,t}$, increases with tree size and canopy exposure (access to sunlight). The probability is parameterized as a logit link to diameter D and canopy exposure λ

$$\theta_{ij,t} = \frac{\exp\left(\beta_0^\theta + \beta_1^\theta D_{ij,t} + \beta_2^\theta \lambda_{ij,t}\right)}{1 + \exp\left(\beta_0^\theta + \beta_1^\theta D_{ij,t} + \beta_2^\theta \lambda_{ij,t}\right)}. \tag{17.1}$$

For computation we require conditional probabilities that derive from this relationship. Two simple examples we detail in the appendix include the probability of making the transition in year t, given previous immaturity and future maturity

$$\delta_{ij,t} = \Pr\left(Q_{ij,t} = 1 \,|\, Q_{ij,t-1} = 0,\ Q_{ij,t+1} = 1\right) = \frac{d\theta_{ij,t}}{d\theta_{ij,t} + d\theta_{ij,t+1}}$$

where

$$d\theta_{ij,t} = \beta_1^\theta d_{ij,t} \theta_{ij,t}\left(1 - \theta_{ij,t}\right) dt$$

and the probability that the transition year τ_{ij} occurred in year t

$$\delta_{ij} = \Pr\left(\tau_{ij} = t \,\middle|\, \tau_{ij}^0 < \tau_{ij} < \tau_{ij}^1\right) = \frac{d\theta_{ij,t}}{\left(\theta_{ij,\tau_{ij}^1} - \theta_{ij,\tau_{ij}^0}\right)}$$

where the lower limit represents that last year the individual was known to be immature and the upper limit represents the first year the individual was known to be mature. These two fundamental relationships are the basis for models that incorporate observations for dioecious and monoecious species, respectively (Appendix: Computation). The first is a Bernoulli probability for each year t; there is a probability for every tree year. The second is a probability associated with an individual and depends on when the maturation event occurred for that individual.

For monoecious species the joint distribution of maturation and fecundity is represented by a discrete mixture that is conditional on the full history of status observations on the individual, a vector $q_{ij} = \{q_{ij,t},\ t = (t_{ij}, \ldots, T_{ij})\}$,

$$\begin{aligned} &p\left(Q_{ij,t},\ f_{ij,t} \,|\, q_{ij},\ x_{ij,t-1},\ Q_{ij,t-1},\ Q_{ij,t+1}\right) \\ &\quad = \left(1 - \delta_{ij,t}\right)^{1 - Q_{ij,t}} \left[\delta_{ij,t}\, N\left(\ln f_{ij,t} \,\middle|\, \mu_{ij,t}^{f|d},\ V_{ij,t}^{f|d}\right)\right]^{Q_{ij,t}} \end{aligned} \tag{17.2}$$

where the probability $\delta_{ij,t} = p(Q_{ij,t} = 1 | q_{ij},\ Q_{ij,t-1} = 0,\ Q_{ij,t+1} = 1)$ is based on the history of observations on ij, and the parameters for the log normal

distribution are conditional, coming from a bivariate state-space model for fecundity (Appendix: Computation).

For dioecious species, we require gender and the probability for the entire history of maturation Q_{ij} and the probability associated with it, δ_{ij}. This distribution is given by

$$p\left(H_{ij},\, Q_{ij},\, f_{ij} \,|q_{ij},\, x_{ij,t-1}\right) = (1-\phi)^{1-H_{ij}}\,\delta_{ij}\left\{\phi\prod_t\left[N\left(\ln f_{ij,t}\,\middle|\,\mu_{ij,t}^{f|d},\, V_{ij,t}^{f|d}\right)\right]^{Q_{ij,t}}\right\}^{H_{ij}}. \qquad (17.3a)$$

If an observation establishes an individual as male we have the probability for a maturation history during which no reproduction occurred

$$p\left(Q_{ij},\, f_{ij}=0 \,|q_{ij},\, x_{ij,t-1},\, H_{ij}=0\right) = \delta_{ij} \qquad (17.3b)$$

and for a female during which reproduction may or may not have occurred

$$p\left(Q_{ij},\, f_{ij} \,|q_{ij},\, x_{ij,t-1},\, H_{ij}=1\right) = \delta_{ij}\prod_t\left[N\left(\ln f_{ij,t}\,\middle|\,\mu_{ij,t}^{f|d},\, V_{ij,t}^{f|d}\right)\right]^{Q_{ij,t}}. \qquad (17.3c)$$

Thus far, we have conditional relationships involving fecundity and maturation. The conditional dependence on seed rain data is discussed in the next section.

17.3.2 Seed data and fecundity

Fecundity (seed production per individual per year) is not directly observed, because seeds cannot be counted in crowded canopies. Like gender and maturation status, indirect information comes from seed trap data, linked by way of a transport model. Individuals that are mature and female can produce seeds. Fecundity is thus zero for immature individuals and all male trees. For mature females fecundity is taken as a continuous, positive variable.

Seeds accumulating in traps located throughout each stand j provide a basis for inverse modeling of fecundity. The likelihood for seeds collected in trap k in stand j in year t is taken to be Poisson

$$Po\left(s_{jk,t}\,|\,A_{jk}g_{jk}\left(f_{j,t};u\right)\right) \qquad (17.4)$$

where A_{jk} is the area of the seed trap ($0.16\,m^2$ or $0.125\,m^2$), and $g_{jk,t}$ is the expected density of seed (m^{-2}), including a parameter u. The expected seed density depends on fecundities of all trees in stand k and a dispersal kernel K, added to a crude estimate of small background density of seed that might enter the plot from outside the mapped boundaries, proportional to basal area of the species in stand j, or $BA_{j,t}$.

$$g_{jk}\left(f_{j,t}\right) = c\cdot BA_{j,t} + \sum_i f_{ij,t}\,K\left(r_{ik};u\right). \qquad (17.5)$$

The dispersal kernel is taken to be a two-dimensional Student's t, previously found to fit seed dispersal data well,

$$K(r_{ik};u) = \frac{1}{\pi u\left(1 + r_{ik}^2/u\right)^2} \tag{17.6}$$

containing the scale parameter u (Clark *et al.* 1999, 2004). The term in equation (17.5) that includes basal area BA allows for the fact that some small fraction of seed can derive from outside the plot, roughly proportional to the basal area of the species.

In addition to seed data, fecundity depends on covariates. In the appendix we discuss conditional relationships involving the likelihood for seed data (equation 17.3) and a multivariate regression for growth and fecundity, influenced by covariates. We include as covariates tree diameter and light availability, summarized by exposed canopy area. The relationship between covariates and fecundity is described by a bivariate state-space model that additionally includes growth (diameter increment) as a response variable. This bivariate model is described with diameter growth in the next section.

17.3.3 Diameter growth and fecundity

Diameter growth (cm per year) is informed by two sparse data sets and by the state-space model that includes fecundity. Censuses conducted at two to four year intervals, which include measurements of diameter on all trees, provide observations of diameter change over the measurement interval. Increment cores are obtained for some individuals and provide annual rates of growth up until the year in which the core was extracted. Because they are laborious to obtain and they can damage trees, increment cores are not available for many trees. Thus, both types of data are sparse, but in different ways; census data exist for all individuals, but only in a few years, and increment cores were taken from a subset of individuals, but cover all years up until the year the core was collected. In consideration of the multiple data types and sparsity of both, diameter growth was modeled in a two-step process, the first step being a model that assimilates the different types of data and generates posterior estimates of growth for each tree year. This analysis is described in Clark *et al.* (2007) and Metcalf *et al.* (2009). Products of this analysis include estimated means and standard deviations for diameter and diameter increment in each tree year (Figure 17.2). These are used as priors for the second step, the analysis described here.

The state-space model for growth and fecundity was developed to allow for measurable covariates known to affect demography, random effects at the individual level, year effects, and error. Random effects are included, because many factors could affect individual health that cannot be assigned to observed

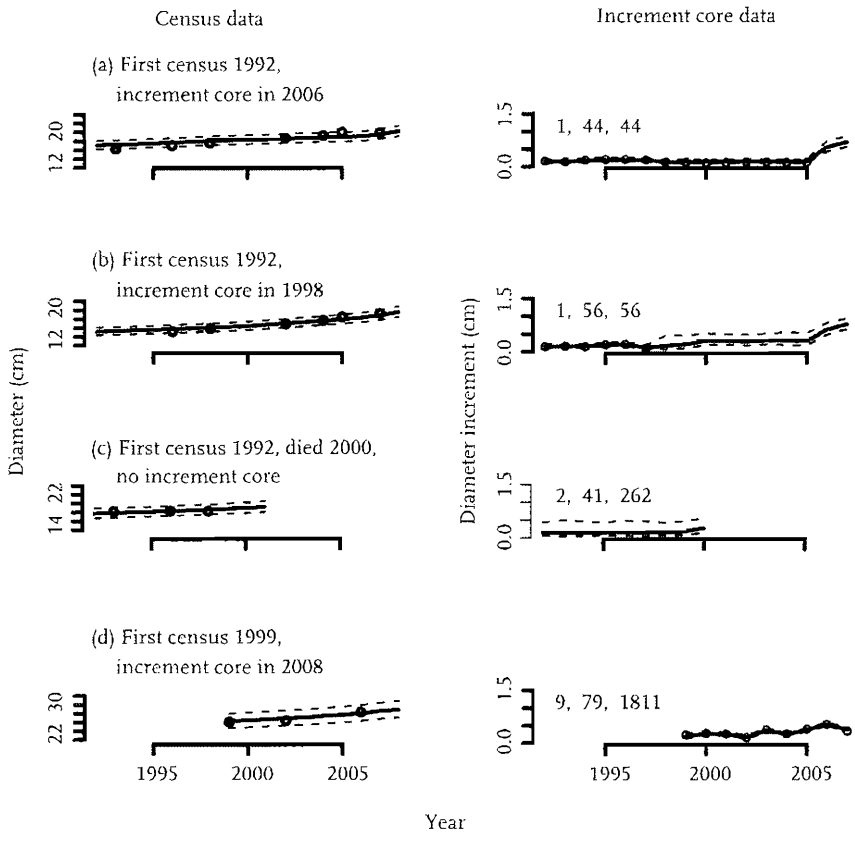

Fig. 17.2 Posteriors for diameter at left and diameter increment at right, represented by median (solid) and 95% CI (dashed) compared with observations (dots). Series vary in length, depending on when observations began and tree survival. Posteriors are generally narrow where increment core data are available, and vice versa. At left are shown corresponding intervals for tree diameter with diameter census data (dots). Numbers at right indicate (individual i, plot j, and a unique order number in the data base). This example is for *Quercus*.

variables. These factors are related to genotype and spatial variation in resources and factors that limit growth. Year effects are included because year-to-year variation in climate is difficult to quantify in ways that might be important for trees. Year effects allow for variation in time that is shared across the entire population. The growth-fecundity submodel is

$$\begin{aligned} y_{ij,t} &= x_{ij,t-1}A + b_t + b_{ij} + \varepsilon_{ij,t} \\ b_{ij} &\sim N_2(0,\ V_b) \\ \varepsilon_{ij,t} &\sim N_2(0,\ \Sigma) \end{aligned} \tag{17.7}$$

with a response vector that includes diameter growth and fecundity

$$y_{ij,t} = \left[\ln\left(d_{ij,t}\right),\ \ln\left(f_{ij,t}\right)\right]; \tag{17.8}$$

covariates

$$x_{ij,t-1} = \left\lfloor 1, \ln\left(D_{ij,t-1}\right), \ln^2\left(D_{ij,t-1}\right), \ln\left(\lambda_{ij,t-1}\right), \ln\left(d_{ij,t-1}\right) \right\rfloor; \tag{17.9}$$

parameters for fixed covariate effects A, fixed year effects b_t, random individual effects b_{ij}, and error $\varepsilon_{ij,t}$. If the genus includes multiple species, the vector $x_{ij,t}$ includes a fixed effect for each species of the genus. Priors for regression parameters are specified such that $\ln(D)$ term takes up allometric effects of size on fecundity when trees are small, and $\ln^2(D)$ takes up senescence effects on growth and fecundity when trees are large. Growth and fecundity rates are both affected by individual size and by abiotic covariates in any given year. By including allometric relationships in the $\ln(D)$ term, the exposed canopy area λ term estimates the effect of light availability (as opposed to size) on growth and fecundity (Section 17.4).

In addition to the regression for growth and fecundity, both variables conditionally depend on data. For fecundity, conditional dependence for $f_{ij,t}$ includes all seed traps on plot j in year t. Because seed trap count $s_{jk,t}$ depends, in turn, on all trees on plot j, the conditional dependence for all trees on plot j in year t is $f_{j,t} > 0$ involves equations (17.2) and (17.3),

$$p\left(f_{j,t} \middle| s_{j,t}, \ldots\right) \propto \prod_k Po\left(s_{jk,t} \middle| A_{jk} g_{jk}\left(f_{j,t}; u\right)\right) \prod_i p\left(f_{ij,t} \middle| Q_{ij,t}, q_{ij}, x_{ij,t-1}\right). \tag{17.10}$$

The second factor on the right hand side comes from equation (17.2). The conditional means and variances for equation (17.2) are

$$\begin{aligned} \mu_{ij,t}^{f|d} &= \mu_f + \frac{\Sigma_{12}\left(\ln d_{ij,t} - \mu_d\right)}{\Sigma_{11}} \\ V_{ij,t}^{f|d} &= \Sigma_{22} - \Sigma_{12}^2/\Sigma_{11} \end{aligned} \tag{17.11}$$

Σ_{ij} is the ijth element of Σ, and unconditional means are

$$\begin{aligned} \mu_d &= x_{ij,t-1} a_{\bullet d} + b_{d,t} + b_{d,ij} \\ \mu_f &= x_{ij,t-1} a_{\bullet f} + b_{f,t} + b_{f,ij}. \end{aligned}$$

The terms on the right hand side include the columns of A, i.e. $A = [a_{\bullet d}\ a_{\bullet f}]$, and the year and individual effects associated with growth and fecundity. For growth, we have the conditional dependence

$$\begin{aligned} p\left(d_{ij,t}\right) \propto\ & N\left(\ln\left(d_{ij,t}\right) \middle| \mu_{d|f},\ V_{d|f}\right) \\ & N\left(\ln\left(d_{ij,t}\right) \middle| \ln\left(d_{ij,t}^{(0)}\right),\ v_{ij,t}\right) I\left(\left(d_{ij,t}^{(0)} - \sqrt{v_{ij,t}}\right) < \ln\left(d_{ij,t}\right)\right. \\ & \left. < \left(d_{ij,t}^{(0)} + \sqrt{v_{ij,t}}\right)\right) Bernoulli\left(z_{ij,t} \middle| \zeta_{ij,t}\right) \end{aligned} \tag{17.12}$$

where $z_{ij,t}$ is the event that the individual survived interval (t_{-1}, t), $d_{ij,t}^{(0)}$ and $v_{ij,t}$ are the prior mean and variance for log growth rate from the analysis of Clark *et al.* (2007)(Figure 17.2), and conditional means and variances are

$$\mu_{j,t}^{d|f} = \mu_d + \frac{\Sigma_{12}\left(f_{ij,t} - \mu_f\right)}{\Sigma_{22}}$$
$$V_{j,t}^{d|f} = \Sigma_{11} - \Sigma_{12}^2/\Sigma_{22}.$$

Note that the prior for the growth data is truncated to a width of two standard deviations. The standard deviations are large for tree years in which there are census data but no increment data (Figure 17.2). For years lacking increment core data, estimates are more heavily influenced by random effects for individuals and fixed effects for years, thus borrowing information from within the individual over time and from year-to-year variation that is shared by the entire population. The last factor in equation (17.12) comes from survival (Section 17.3.5). The survival probability enters the conditional probability, because it depends on growth rate, through the binned diameter classes (Section 17.3.5).

In addition to equation (17.7) that applies to mature individuals, we fitted an additional model for growth using all trees, regardless of maturation status. This model is univariate, but includes the same covariates as used for mature trees,

$$\begin{aligned} \ln d_{ij,t} &= x_{ij,t-1}a + b_t + b_{ij} + \varepsilon_{ij,t} \\ b_{ij} &\sim N\left(0, V_{b(1,1)}\right) \\ \varepsilon_{ij,t} &\sim N(0, w) \end{aligned} \tag{17.13}$$

where a is the parameter vector (the first column of A in equation 17.7), b_t is the first element of b_t in equation (17.7), and $V_{b(1,1)}$ is the random effects variance for diameter growth (element 1,1 of covariance matrix V_b) from equation (17.7).

17.3.4 Exposed canopy area

Light availability is included as a predictor of growth and fecundity. It is summarized by an index, the area of the crown potentially exposed to sunlight, $\lambda_{ij,t}$. This index has non-zero values, but can be small, such as for the case of a suppressed individual in the understory. It is estimated on the basis of three sources of information in a separate analysis (Wolosin *et al.*, in review). These data sources are (i) low-altitude imagery on which crown area can be measured, (ii) ordinal status classes assigned on the basis of ground observations, and allometric measurements combined with models of solar geometry that yield calculations of light availability throughout the day and over the growing season. A model combines these sources of information. Posterior means and variances enter this analysis as priors, just as described for diameter growth (Section 17.3.3).

17.3.5 Survival

Survival probability is typically modeled as a function of growth rate (see the appendix), which integrates many aspects of tree health (Kobe *et al.* 1995, Clark and Clark 1996, Wyckoff and Clark 2000, Bigler and Bugmann 2004) and of size (Clark and Clark 1996, King *et al.* 2006, Coomes and Allen 2007). A number of functional forms have been used to relate survival to growth rate. The problem with any functional form comes from the facts that (i) this relationship can be strongly nonlinear, changing abruptly at growth rate values close to the lowest range of values typically observed, and (ii) the distribution of data has a large impact on estimates, and individuals close to death may be poorly represented in data sets, because such individuals died disproportionately before the study began. We have developed or modified nonparametric approaches (Wyckoff *et al.* 2000, 2000, Clark *et al.* 2007, Metcalf *et al.*, in review) to describe this relationship and apply one that combines not only growth rate, but also tree size effects on mortality risk (Clark *et al.*, 2007). We include tree size as a predictor of survival, because mortality risk may increase as trees become large and senesce or become susceptible to high winds (Batista *et al.* 1998, Uriarte *et al.* 2004, Rich *et al.* 2007).

Let $z_{ij,t}$ be the event that an individual ij is alive in year t, in which case it survived from year t to $t + 1$ with probability

$$\zeta_{ij,t} = 1 - \left(\mu_{d_{ij,t-1}} + \mu_{D_{ij,t-1}} - \mu_{d_{ij,t-1}}\mu_{D_{ij,t-1}}\right). \tag{17.14}$$

There are discrete bins for both growth rate and diameter. In year $t - 1$ individual ij's growth rate bin is indicated by $\mu_{d_{ij,t-1}}$ and its diameter bin is indicated by $\mu_{D_{ij,t-1}}$. There are monotonicity priors for both sequences, decreasing for growth rate and increasing for diameter (Section 17.4). The likelihood is $Bernoulli(z_{ij,t}|\zeta_{ij,t})$. In the next section we specify priors.

17.4 Prior distributions

The model includes both informative and non-informative prior distributions. Due to the size and complexity of the model, where possible, we used informative priors that are flat but truncated in some fashion, to maximize transparency, i.e. for identification of the contributions of prior versus likelihood. Here we summarize priors and how they were selected to balance information.

The fixed effects in the state-space model (17.7) have flat priors bounded by values either having theoretical justification or sufficiently wide to not impact estimates,

$$vec(A) \sim I(a_1 < vec(A) < a_2) \tag{17.15}$$

where a_1 and a_2 are vectors of minimum and maximum values, respectively. We describe prior values for specific elements of A, using indexing for elements of A that assume a single intercept. Recall that there are separate intercepts for each species included in a given genus (Table 17.3). The actual number of rows of A is p = number of species + 4 (there are four covariates). The first subscript indicates the covariate (equation 17.9) and the second subscript indicates the response (equation 17.8). Truncation points that affect estimates include:

A_{21} – *diameter effect on diameter growth increment constrained to be near zero*: Diameter is included as a predictor, because we expect it to directly affect fecundity – large trees are capable of high seed production. We expect it to also be correlated with the other response variable, diameter growth increment. Because that correlation should be taken up by canopy exposure, rather than by tree diameter directly, we constrain this parameter value to be near zero. From open-grown trees there is no clear evidence for a direct size effect on diameter increment. The correlation between tree size and diameter growth increment is expected to result from the fact that large trees are more likely to have higher light exposure. Because there is no theoretical justification for non-zero values, limits on this parameter are $(-0.02, 0.02)$.

A_{22} – *diameter effect on fecundity constrained between 1.5 and 2.5*: Allometric arguments and empirical evidence suggest that potential fruiting yield should scale with canopy width, which, in turn is roughly proportional to diameter. In fact, this potential should not be realized for trees crowded by neighbours. We allow for this effect of size on potential yield effect due to size with the constraints (1.5, 2.5), as modified by competition, which is reflected in exposed canopy area, i.e. the term including λ.

(A_{31}, A_{32}) – *large diameter effect is negative*: The squared diameter term in equation (17.9) is included to allow for potential senescence, a decline in physiological function with age. Tree data sets rarely have sufficient large (potentially old) individuals to estimate these effects, but we can allow that senescence does eventually occur by specifying that this effect only has impact for especially large individuals. This term is constrained to be negative.

(A_{41}, A_{42}) – *lag-1 effect of growth rate on growth rate and fecundity*: This effect was constrained to be effectively zero for growth rate (A_{41}) but unconstrained for fecundity (A_{42}). We wanted to parametrize the effect on fecundity, so it could be used for predictive modeling of potential tradeoffs in time between growth and fecundity.

The estimates of growth and fecundity represent a balance between contributions from the regression (i.e. the size and light covariates in equation 17.7)

and data models for growth rates and seed data. It is not necessarily 'objective' to use a non-informative prior for the state-space error covariance matrix, because there is no objective criterion for balancing information that enters from multiple data types. Data are known to be noisy, particularly seed rain. We used an informative prior on the error covariance matrix Σ to represent a level of variation expected after that taken up by covariates, random effects, and year effects and to assure that covariates were not overwhelmed by noise. The values used for variances on the log growth (cm) and log fecundity (seeds per tree) were 0.05 and 0.2, respectively. These allow for realistic levels of variation on the non-log scale. We used the prior

$$Wishart\left(\Sigma^{-1} \left| \begin{bmatrix} 0.05 & 0 \\ 0 & 0.2 \end{bmatrix}^{-1}, \quad n_{IJT}\right.\right) \tag{17.16}$$

where $n_{IJT} = \sum_{i,j} (T_{ij} - t_{ij})$ is the number of tree-years in the study. Through extensive sensitivity analysis, this prior was found to provide an acceptable balance of data and regression model, contributing to the conditional posterior approximately twice the weight coming from the regression.

By contrast, priors on random effects and year effects were weak – we wanted data to dominate these estimates. The prior for random effects is

$$Wishart\left(V_b^{-1} \left| \begin{bmatrix} 0.2 & 0 \\ 0 & 2 \end{bmatrix}^{-1}, \quad \max\left(3, n_{IJ}/100\right)\right.\right) \tag{17.17}$$

where $n_{IJ} = \sum_j I_j$ is the number of trees. The second parameter in equation (17.17) is rounded to an integer value and ranged from 3 to 60 for different species. The contribution to the conditional posterior ranged from about 1/10 to 1/50 of the weight coming from the regression.

The prior for fixed year effects is

$$N_2\left(b_t \left| \begin{bmatrix} 0 & 0 \end{bmatrix}', diag(100, 100)\right.\right) \tag{17.18}$$

and includes a sum-to-zero constraint (intercepts are included in A), implemented directly in the Gibbs sampler.

Because many individuals are not mature, a separate univariate regression is fitted to all tree years, regardless of maturation status (equation 17.13). The covariates are the same as those listed for the multivariate regression given above, and sampling makes use of the univariate distributions corresponding to each of the foregoing multivariate ones. These are Gaussian for fixed (including years) and random effects, and inverse Gamma for variances.

Diameter growth increments have a prior for each tree year taken from the posterior from the analysis of Clark *et al.* (2007) and shown in equation (17.12). Because there are thousands of such densities, the truncation of this posterior at a width of two standard deviations was based on a prior belief that true increments should be within this range. Diagnostics showed that posterior estimates from this analysis did not tend to acccumulate at these truncation values.

As with diameter increment, the Gaussian prior for canopy values is truncated to two standard deviations in width,

$$\ln \lambda_{ij,t} \mid \sim N\left(c_{ij,t}^{(o)}, C_{ij,t}\right) I\left(\left(c_{ij,t}^{(0)} - \sqrt{C_{ij,t}}\right) < \ln\left(\lambda_{ij,t}\right) < \left(c_{ij,t}^{(0)} + \sqrt{C_{ij,t}}\right)\right) \quad (17.19)$$

where $c_{ij,t}^{(0)}$ and $C_{ij,t}$ are the prior mean and variance (log scale), taken from the posterior for the analysis of canopy area (Wolosin *et al.*, in review).

Priors for fecundity, maturation, gender, and missing seed data were either non-informative or derived from previous observation. A flat prior was used for fecundity, truncated at the smallest number of seeds observed for a tree and at values much larger than implied by observation of seed densities,

$$f_{ij,t} \sim unif\left(f_{\min}, f_{\max}\right). \quad (17.20)$$

For instance, when defining $f_{\min}$, we did not expect that a mature individual would produce less seeds than typically contained in a single fruiting structure (e.g. *Pinus, Liriodendron, Liquidambar*). For maximum values, we used parameter estimates similar to those obtained in a simpler model (Clark *et al.* 1999) to 'invert' seed density observations and thus approximate what might constitute unrealistically high seed production for an individual of a given species and size. The model for seed dispersal provides an expected seed density given the spatial locations of trees of different sizes. For example, *Acer rubrum* seeds have been observed at average densities of 10^2 seeds m^{-2} beneath mature trees but not at average densities of 10^3 seeds m^{-2}. This inversion was used not only to set limits on fecundities for individual trees, but also to define the (prior) Poisson means for missing seed data.

The maturation diameter for an individual was assigned a prior that was truncated at values below which we believed that no individuals could be mature and above which we believed all individuals would be mature. These beliefs came from independent observations of trees in similar settings. These limits on maturation diameters translate to limits on maturation year (see appendix). There are prior minimum and maximum diameters, which differ among species. The female fraction was given the prior $Be(\phi|h_1, h_2)$ with $h_1 = h_2 = 4$, having a mean of 0.5 and being dominated by the data. The probability of recognizing a mature individual was assigned the prior $Be(v|v_1, v_2)$ with $v_1 = v_2 = 0.002n_{IJ}$, which has a mean of 0.5 and is dominated by the data.

Parameters for the logit function of maturation equation (17.1) were assigned the prior

$$N(\beta^\theta \mid b^\theta, V^\theta)\, I(\beta^\theta_{2,3} > 0) \tag{17.21}$$

a truncated normal prior with mean vector $b^\theta = [-3,\ 0.1,\ 0.1]'$ and covariance $V^\theta = diag[10,\ 10,\ 10]$. The positivity constraint on the second and third elements of the vector comes from the prior belief that the relationship between maturation and diameter and light availability is non-negative.

Priors for the seed data model in equation (17.3), including the dispersal parameter and the seed fraction originating outside the map, were

$$p(u, c) = N(u \mid u_0, V_u)\, N(c \mid c_0, V_c)\, I(u, c > 0) \tag{17.22}$$

where parameter values were chosen to be informative. For u they differ among species; we used $c_0 = 0.02$ and $V_c = 0.01$. There is a positivity constraint on u and c.

The monotonicity priors on the parameter sequences μ_d and μ_D in equation (17.14) were designed to allow for uneven distribution of data and strong nonlinearities. Because slow growth is associated with death, the observations of growth rate below a certain threshold are rarely observed. However, this lack of slow growth observations results from the fact that mortality risk increases sharply at slow growth rates. For this reason, our sequence of μ_d values has an intercept at 1. Although zero growth rates do occur in particular years, we used this assumption as a way of approximating the sharp increase in risk that can occur at low growth. This assumption is obviously flexible. In addition to monotonicity, there was an informative prior for values within the sequence μ_D, which was $Be(a_k, b_k)$, where a_k is 0.001 for $k = 1, 2, 3$ and $a_k = 10$ for $k = 4, 5, 6$,

$$b_k = a_k \left(\frac{1}{\mu_k^0} - 1 \right)$$

$$\mu_k^0 = [0.00001,\ 0.00002,\ 0.00003,\ 0.00004,\ 0.00005]. \tag{17.23}$$

This prior assures essentially zero values for juvenile trees (bins 1, 2, 3), but is non-informative (but monotonically increasing) for large trees. Thus, the diameter effect only affects large trees. Although small trees grow slowly and thus are at higher mortality risk, this is a growth effect, not a diameter effect. This informative prior allows us to separate the effect of slow growth from that of large size, which could indicate senescence.

17.5 Computation

The posterior was simulated with Gibbs sampling, based on conditional posteriors that are discussed in the appendix, some of the embedded steps being

Metropolis. The simulation was initialized at prior mean parameter values (diameter increments and crown areas), random draws from priors, or MLEs based on simpler models (fecundities for trees were initially estimated without year or individual effects using the approach of Clark *et al.* 1999).

Due to the size of the model, efforts were made to optimize code. Despite the large number of years across many individuals within multiple plots, the main Gibbs loop contains only three loops over years (including one to update maturation/fecundity, one for missing seed data, and another for dispersal and Poisson parameters), and no loops over individuals or plots. Data structures that include pointer arrays were used to rebuild (reorder and restack) matrices of state variables based on the changing gender and maturation statuses of trees and tree-years, respectively.

Convergence was achieved with 10,000 iterations for species with moderate numbers of individuals, but required up to 200,000 iterations for trees with many individuals. There are a large number of parameters, not all of which could be sampled efficiently. The lowest updating rates and highest autocorrelations were obtained for fecundities of dioecious species (*Acer rubrum, A. pennsylvatica, Fraxinus americana,* and *Nyssa sylvatica*), due to the discrete nature of $Q_{ij,t}$ and H_{ij}, and the blocking over all tree-years within a plot. Thus, for fecundity/maturation/gender of dioecious species, we selected for updating at random 30% of the trees for a given iteration and embedded five such iterations within each Gibbs step.

17.6 Diagnostics

From 50,000 to 1,000,000 Gibbs steps were discarded, followed by 50,000 to 100,000 iterations that were retained for analysis. We inspected Gibbs chains for all parameters as well as for samples of individual effects. Experiments involved many parameter initializations; however, results presented here come from single long runs for each taxon group. Acceptance rates for Metropolis steps were generally above 0.2. The exception was for dioecious species, where low acceptance rates were addressed by embedding multiple iterations per Gibbs step (Section 17.5). To help evaluate results we compared priors and posteriors, we considered predictive capacity, in terms of data used to the fit model, and we compared predictive intervals from the model with estimates of latent states that could not be directly observed. Here we discuss some comparisons.

17.6.1 Some prior/posterior comparisons

Some of the estimates for parameters from an example taxon, *Quercus*, are shown in Table 17.5. Estimates for marginal posteriors are accompanied by fitted truncated normal distributions, which would be used in the event that it

Table 17.5 Posterior estimates for some of the main parameters in the model, shown for the taxon *Quercus*.

Symbol (equation)	Parameter p[1]	Marginal posterior				Fitted model[2] $N(p\|m, s^2)I(p_1 < p < p_2)$			
		mean	std dev	0.025	0.975	m	s	p_1	p_2
A (17.7)	quun.d	−1.65	0.0271	−1.7	−1.59	−1.65	0.0271	−4	1
	quun.f	4.15	0.283	3.7	4.65	4.15	0.283	−1	10
	diam.d	0.0196	0.000394	0.0186	0.02	0.393	0.0122	−0.02	0.02
	diam^2.d	−0.000155	5.55e − 05	−0.000301	−0.000101	0.00662	0.000619	−0.2	−1e−04
	cnpy.d	0.0591	0.00478	0.0496	0.0683	0.0591	0.00479	0.01	1
	dlast.d	0.00992	8.34e − 05	0.00969	0.01	0.0376	0.00163	−0.01	0.01
	qual.d	−1.3	0.0151	−1.33	−1.27	−1.3	0.0151	−4	1
	quco.d	−1.66	0.0168	−1.69	−1.62	−1.66	0.0169	−4	1
	qufa.d	−1.26	0.0324	−1.32	−1.2	−1.26	0.0324	−4	1
	quma.d	−1.61	0.0359	−1.68	−1.54	−1.61	0.0359	−4	1
	quph.d	−1.37	0.0182	−1.41	−1.34	−1.37	0.0182	−4	1
	qupr.d	−1.7	0.0115	−1.72	−1.68	−1.7	0.0115	−4	1
	quru.d	−1.62	0.0123	−1.64	−1.6	−1.62	0.0123	−4	1
	qust.d	−1.5	0.0214	−1.54	−1.46	−1.5	0.0214	−4	1
	quve.d	−1.76	0.0219	−1.8	−1.72	−1.76	0.022	−4	1
	diam.f	1.51	0.0117	1.5	1.54	0.473	0.0724	1.5	2.5
	diam^2.f	−0.183	0.00782	−0.193	−0.159	−0.183	0.00782	−0.25	−1e−04
	cnpy.f	0.131	0.0277	0.0777	0.183	.131	0.0278	0.01	3
	dlast.f	−0.0891	0.0949	−0.236	0.107	−0.0891	0.095	−2	2
	qual.f	4.09	0.16	3.84	4.51	4.09	0.161	−1	10
	quco.f	3.9	0.186	3.6	4.32	3.9	0.186	−1	10
	qufa.f	4.03	0.202	3.72	4.54	4.03	0.202	−1	10
	quma.f	4.06	0.181	3.74	4.45	4.06	0.181	−1	10
	quph.f	4.29	0.155	4.06	4.68	4.29	0.155	−1	10
	qupr.f	3.9	0.18	3.63	4.34	3.9	0.18	−1	10
	quru.f	4.28	0.159	4.03	4.69	4.28	0.16	−1	10
	qust.f	4.19	0.151	3.95	4.58	4.19	0.151	−1	10
	quve.f	4.02	0.187	3.75	4.52	4.02	0.187	−1	10

a (17.13)	diam	0.0185	0.00137	0.015	0.02	0.184	0.016	−0.02	0.02
	$diam^2$	−0.0538	0.00564	−0.0669	−0.0434	−0.0538	0.00564	−1	−0.001
	cnpy	0.0228	0.00366	0.0155	0.03	0.0228	0.00366	0.005	1
	dlast	0.00987	0.000129	0.00952	0.01	0.0394	0.00211	−0.01	0.01
	qual	−0.76	0.18	−1.1	−0.398	−0.761	0.18	−4	0
	quco	−1.08	0.289	−1.6	−0.477	−1.08	0.289	−4	0
	qufa	−0.554	0.354	−1.36	−0.0301	−0.397	0.465	−4	0
	quma	−0.95	0.464	−1.97	−0.107	−0.907	0.509	−4	0
	quph	−0.67	0.275	−1.21	−0.165	−0.666	0.282	−4	0
	qupr	−1.14	0.0674	−1.27	−1.01	−1.14	0.0675	−4	0
	quru	−1.08	0.114	−1.31	−0.884	−1.08	0.114	−4	0
	qust	−0.736	0.345	−1.44	−0.149	−0.728	0.357	−4	0
	quve	−1.24	0.272	−1.67	−0.57	−1.24	0.272	−4	0
	quun	−1.12	0.471	−2.2	−0.305	−1.12	0.473	−4	0
Σ (17.7)	$\Sigma_{(1,1)}$	0.0511	0.000422	0.0503	0.0519	–	–	–	–
	$\Sigma_{(2,2)}$	0.218	0.00174	0.214	0.221	–	–	–	–
	$\Sigma_{(1,2)}$	$6.35e-05$	0.000593	−0.00111	0.00123	–	–	–	–
β^r (17.4)	intercept	0.757	0.102	0.539	0.935	–	–	–	–
	slope	0.1	0.00714	0.0887	0.118	–	–	–	–
V_b (17.7)	$V_{b\ (1,1)}$	0.105	0.00433	0.0967	0.114	–	–	–	–
	$V_{b\ (2,2)}$	0.34	0.0177	0.307	0.375	–	–	–	–
	$V_{b\ (1,2)}$	0.0205	0.00977	0.000808	0.0387	–	–	–	–
β^{fl} (17.2)	intercept	−1.42	0.0421	−1.5	−1.34	–	–	–	–
	diameter	0.0891	0.00296	0.0826	0.0942	–	–	–	–
	light	0.0197	0.00355	0.0126	0.0264	–	–	–	–
w (17.13)	error variance	0.0415	0.00054	0.0404	0.0425	–	–	–	–
u (17.6)	dispersal	29.9	0.0624	29.8	30	–	–	–	–
v	detection	0.0732	0.00371	0.0662	0.0807	–	–	–	–

[1] For *A*, parameters are referenced as 'covariate.d' or 'covariate.f'. Thus, 'cnpy.d' is the effect of canopy, or light response, on diameter growth d, and 'quru.f' is the fecundity intercept for *Quercus rubra*. For *a*, there is a single response variable.

[2] p_1 and p_2 are the lower and upper limits for the truncated normal. For *A*, they are referenced in equation (17.15).

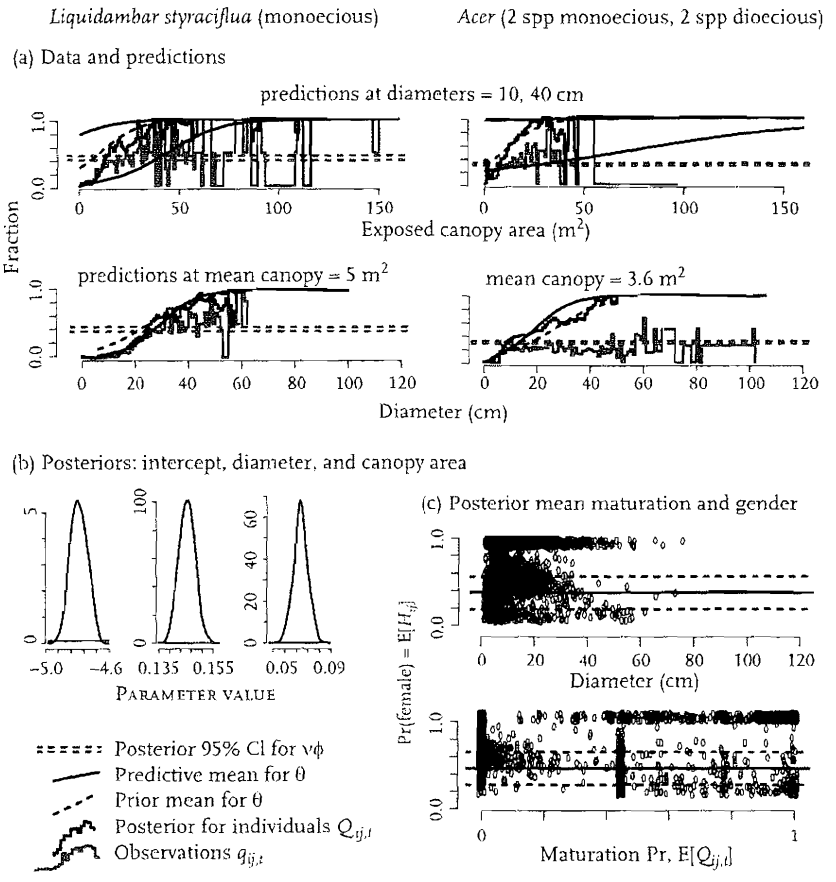

Fig. 17.3 (a) Data and fitted models for maturation and gender. *Liquidambar* (left) is monoecious, *Acer* (right) includes both monoecious and dioecious species. Lower histograms show the faction of observations in diameter bins for which $q_{ij,t} = 1$. Upper histograms show the fraction for which the estimates $Q_{ij,t} = 1$. The function θ is shown for prior (dashed) and posterior (solid) values of β^{θ}. Horizontal dashed lines are 95% CIs for $\nu\phi$. (b) Posteriors for β^{θ} for *Liquidambar* compared with priors (flat lines). (c) Posterior means for gender plotted against maturation status.

was desirable to draw samples from it, for purposes of prediction. This would be necessary if one did not have access to the full Gibbs chains. There are two such distributions, one for the parameters of A and second for those of a. The full covariances are not included for space considerations; we have included only standard deviations in Table 17.5. We consider aspects of data, priors, and posteriors, beginning with gender and maturation, followed by growth/fecundity, then survival.

Figure 17.3 provides perspective on how data, priors, and posteriors relate for a monoecious species (*Liquidambar styraciflua* on the left side of Figure 17.3) and the genus *Acer*, which includes both monoecious (*A. barbatum, A. saccharum*) and dioecious (*A. rubrum, A. pensylvanicum*) species (right side of

Figure 17.3). In fact, *A. rubrum* is polygamo-dioecious, having some individuals that are male, some female, and some supporting both male and female flowers. Our female fraction for this species includes both female and monoecious individuals. For the monoecious *Liquidambar*, all mature individuals have female function, so $\phi = 1$. With increasing diameter and canopy area, larger numbers of individuals are observed to be mature (grey histogram in Figure 17.3a) and still more are estimated to be mature (black histogram), because detection probability $v < 1$ (horizontal dashed grey lines). Note that the posterior 95% credible interval for the estimate of v roughly averages the red histogram (observations) at sizes where maturation is reached, whereas the black histogram (estimates) approaches 1. This is the expected relationship between observations, detection probability, and the true states. The values approach zero for small diameters, because small trees cannot reproduce. However, values do not approach zero for small exposed canopy areas, because it is possible for trees that are highly shaded to produce at least some fruit.

The estimates for the population-level relationship are given by predictions of θ, shown in Figure 17.3 as predictive means only. These are plotted against exposed canopy area λ (for two values of diameter) and against diameter D (for the mean canopy area). We show prior and posterior means for θ. The estimates of β^{θ}, which are the basis for predictions of θ, are well resolved (Figure 17.3b). They predict maturation at smaller sizes and at lower canopy exposure for *Acer* than they do for *Liquidambar*. The population level predictions (smooth lines) do not appear to run precisely through the histograms of individual level predictions, because the individual level predictions effectively marginalize over diameter and canopy distributions for the entire population, whereas the predictive mean curves are conditional on specific diameter and canopy values.

Capacity to predict gender increases with tree size, because large trees are more likely to be reproductive, and reproduction is the only evidence for gender. The probability of being female tends to zero or one with increasing diameter (Figure 17.3c). As probability of being mature increases, so too does predictability of gender. If data were static, at small diameters, the probability that any individual is female would tend to the posterior estimate of ϕ. This does not occur in Figure 17.3c, because small diameter individuals may later become mature, thus providing evidence of their gender even at small size, i.e. before they were mature. With increasing confidence in maturation status, we see a greater tendency to be female than male. This tendency results from the fact that two of the species in *Acer* are modeled as monoecious and thus will always be counted in the female class.

The influence of truncated priors (equation 17.15) is evident in posteriors for parameters from the growth/fecundity state-space model (Figure 17.4). Due to the size and complexity of the model the flexibility to assign hard boundaries to one or both limits for these parameters and the transparency of prior effects

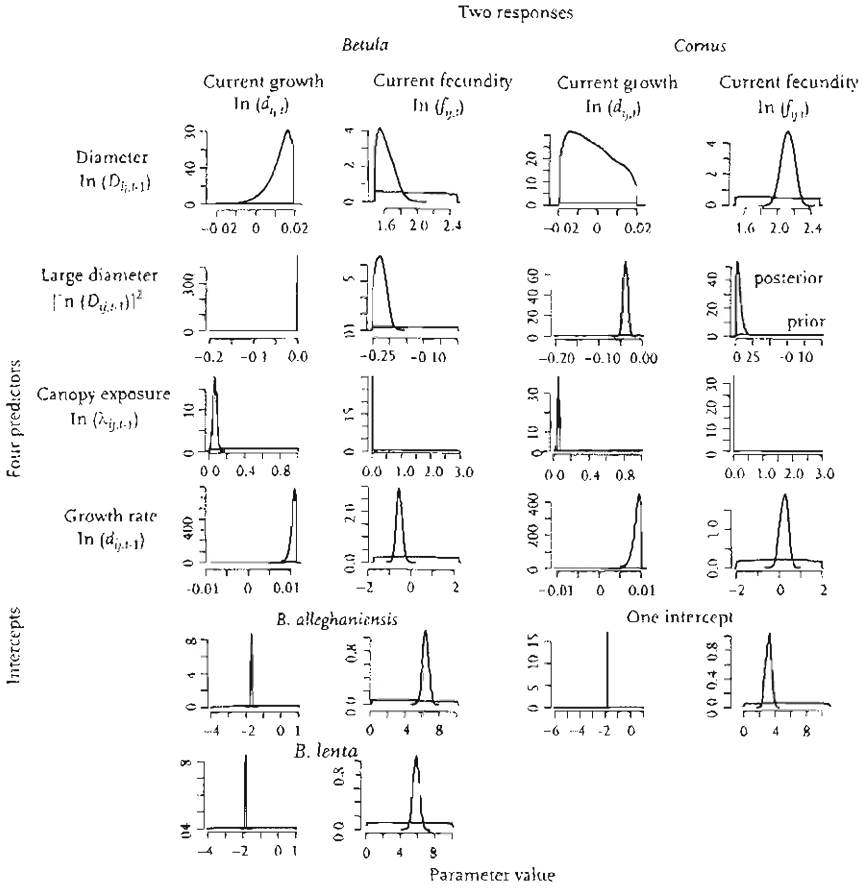

Fig. 17.4 Comparison of priors (flat grey lines) and posteriors (black) for the fixed effects in the matrix A for the state space model (Equation 17.7) for two genera, *Betula* and *Cornus*. *Betula* has two intercepts, one for each of two species. The horizontal axis bounds the prior.

on posteriors was deemed an advantage. We include in this example two shade-tolerant taxa (the canopy exposure effects λ are near zero for both diameter growth and fecundity), one having high fecundity (*Betula*) and another low fecundity (*Cornus*). The fit for *Betula* includes two sets of intercepts, one for the higher fecundity and faster growing *B. alleghaniensis* and one for the lower fecundity and slower growing *B. lenta* (lower panels of Figure 17.4). For this particular fit, we held the diameter effect D on growth rate d to be near zero (there is no prior knowledge to suggest growth rate should respond directly to size until trees become large), but assumed that the effect of D on fecundity should fall between 1.5 and 2.5. Together these assumptions allow for a direct size effect on fecundity that accords with allometric theory, thus allowing that effects of canopy exposure, which can be correlated with size, are more realistic.

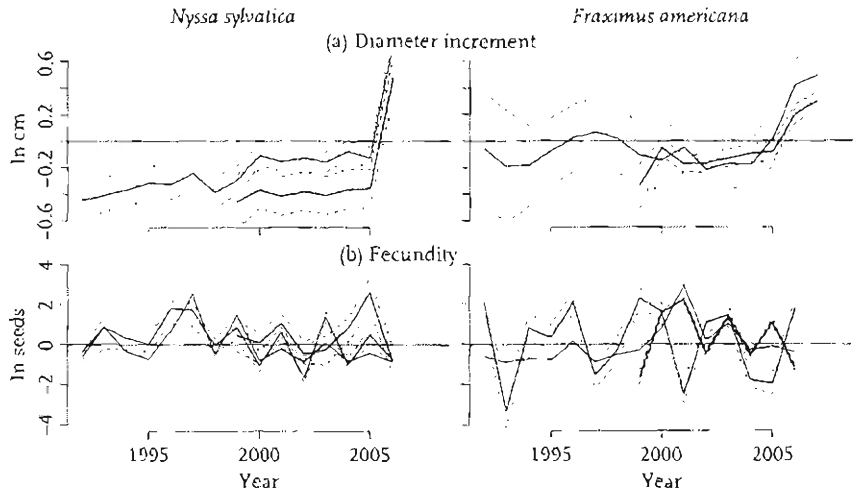

Fig. 17.5 Posterior medians and 95% CIs for year effects b_t for two species. Separate year effects were used for southern Appalachian plots and for Piedmont plots. Solid thin lines in lower panels are proportional to log seed rain averaged over both sites.

To allow for declining physiological function with size, we included the $(\ln D)^2$ term in the model and constrained it to be negative. This term will have increasing influence with size. We did not have large enough trees in these data sets to show clear effects on growth (posteriors clumped at the upper zero boundary), but there was evidence for this negative effect on fecundity for a number of species. Canopy exposure λ has a positive effect on both growth and fecundity. For these shade tolerant species, these estimates were close to zero for both growth and fecundity.

The lagged growth rate effect was constrained to be near zero for growth, because we wanted long-term trends in growth to be taken up by year effects. The tendency for positive correlation was constrained by the upper boundary at 0.01. However, we wanted to explicitly parameterize the lag-1 effect of growth on fecundity, because this could be important for demographic prediction. We obtained a range of values from strongly positive to strongly negative for the lag-1 effect of growth on fecundity.

The different intercepts for genera having more than one species (e.g. *Betula*) allowed us to model fecundities for groups of species having indistinct seeds. Although *B. alleghaniensis* and *B. lenta* have similar life histories, we found substantially higher fecundity for *B. alleghaniensis* (recall that intercepts are on a log scale).

Two sets of fixed year effects b_t were used, one for each of the two regions and shown in Figure 17.5 as longer curves for Coweeta (back to 1992) and shorter curves for Duke Forest (beginning in 1999), both having a sum to zero

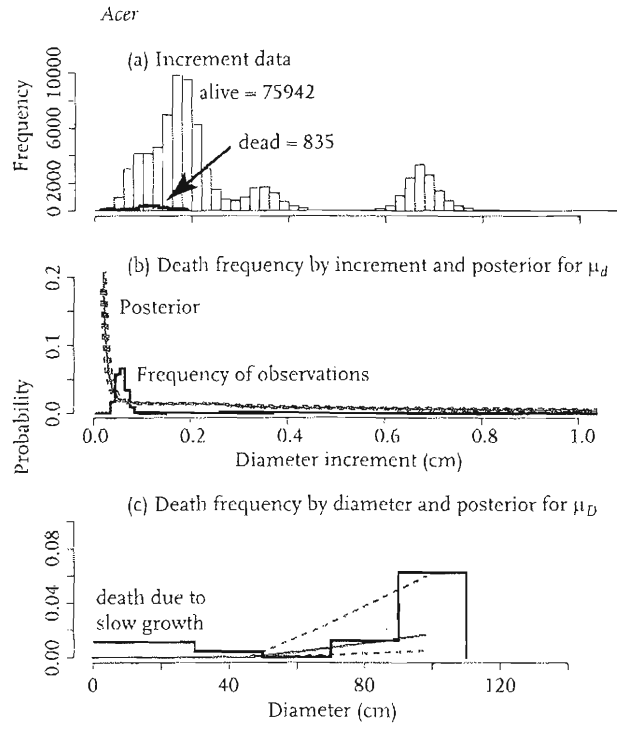

Fig. 17.6 (a) Survival data, (b) relative frequency of deaths (histogram) and posterior median and 95% CI for μ_d, (c) relative frequency of deaths (histogram) and posterior median and 95% CI for μ_D.

constraint. Those for growth show increasing rates in recent years, which could result from several mechanisms. Those for fecundity show a tendency for two-year cycles in *Nyssa*. Note that year effects need not strictly track seed rain trends (solid thin line in Figure 17.5), because other covariates vary from year to year.

The posterior estimates for effects of diameter increment μ_d and diameter μ_D on survival show the effects of the monotonicity assumptions (Figure 17.6). The relationship between growth increment and mortality risk is highly nonlinear close to the limit of increment core data (Figure 17.6a, b). Apparently, trees reach a threshold of low growth, below which mortality risk rises substantially. The histogram of observations in Figure 17.6(c) shows modes not only at the largest sizes, but also the smallest. The latter mode results from the slow growth that results from low light levels in the forest understory. The priors help to discriminate the growth from size effects, by recognizing that mortality risk declines with growth rate, but increases with size. Beyond this relationship already known from previous studies, the prior is weak as to the shapes of these relationships. The prior from equation 17.23 allows the assumption that death

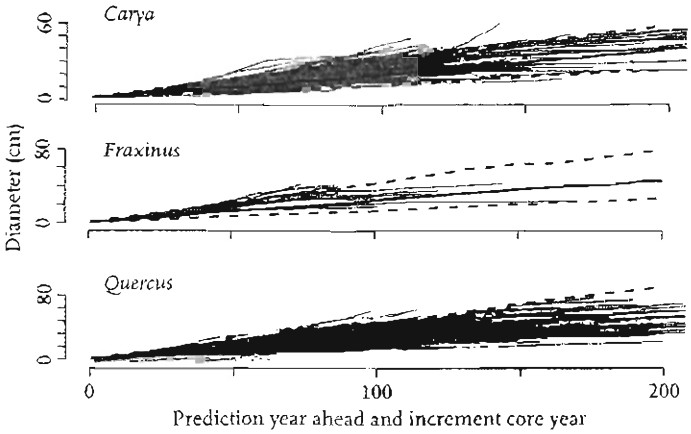

Fig. 17.7 Comparison of increment data from tree ring data not used in fitting the model (thin lines) and predictive distributions of tree diameter from the model (Section 17.6.2).

of these small individuals is not due directly to size, but rather indirectly, due to low growth rates.

17.6.2 Data prediction

To provide further insight into model behaviour we predicted data and latent states. Here we describe some of these predictions and how they compare to data or estimates of latent states.

Predictions of diameter growth were evaluated against an independent data set of growth, obtained from measurements of increment cores spanning decades. The example in Figure 17.7 is typical – we obtain good coverage of size distributions for decade-ahead prediction. It is worth mention that the model contains no explicit age information. And there is no attractor in the model that would necessarily make it converge to a particular diameter value. Moreover, these are not one-step ahead predictions, as is often used to evaluate fits of time series models, but rather 200 year ahead predictions. Here we simply initialized the model and incremented year-by-year predictive distributions, approximating

$$p(d_{t+1}, f_{t+1} | X_t^*, X) = \int p(\gamma_{t+1} | X_t^*, \pi)\, p(\pi | X) d\pi$$

where X is taken to be all data and priors entering the model, π is a vector of parameters, X_t^* is taken to be the current covariates, which include the previously predicted diameter and increment, and a random draw from the distribution of canopy exposure values $\lambda_{ij,t}$ contained in the data. The integrand includes the state-space structure of the model (equation 17.7) and the posterior.

The integral is approximated by drawing at random a row from the iteration-by-parameter matrix of Gibbs sampler output. The tree is initially immature (1 cm diameter) and subject to growth rate in equation (17.13), risk of maturation from equation (17.2), and risk of death from equation (17.14). A draw from V_b determines its random individual effect. If it does not survive an iteration, it is pronounced dead and removed from the simulation. The next growth rate is drawn from a univariate or bivariate normal, depending on maturation status.

A similar approach to prediction was used to evaluate other aspects of the model. We do not observe fecundity, so we cannot compare direct observations of fecundity against model predictions. However, we can check predictions of seed rain. We did this in two ways. Consider that seed rain can be predicted from different levels in the model. The model generates estimates of latent states $f_{ij,t}$. Based on these latent states for all trees on plot j in year t, there is a likelihood for seed rain data at location k in year t. Thus we can consider how well the expected seed rain for all trees at j in year t predict seed rain data at k in year t, or

$$p\left(s_{k,t} \mid E\left[f_{j,t}, Q_{j,t}\right], X\right) = \int p\left(s_{k,t} \mid E\left[f_{j,t}, Q_{j,t}\right], \pi_1\right) p\left(\pi_1 \mid X\right) d\pi_1$$

where X represents all data and priors, and the vector $\pi_1 = (u, \beta^\gamma)$. Alternatively, we could predict from a lower level to include the uncertainty in $f_{j,t}$ itself

$$p\left(s_{k,t} \mid E\left[X_{j,t-1}\right], X\right) = \int p\left(s_{k,t} \mid f_{j,t}, Q_{j,t}, \pi_1\right) p\left(f_{j,t}, Q_{j,t} \mid E\left[X_{j,t-1}\right], \pi_2\right) \times p\left(f_{j,t}, Q_{j,t}, \pi_1, \pi_2 \mid X\right) d\left(f_{j,t}, Q_{j,t}, \pi_1, \pi_2\right)$$

where the vector $\pi_2 = (\beta^\theta, A, a, b_t, \{b_{ij}\}, \Sigma)$. Figure 17.8 compares predictions from these two levels with data (black) and priors for missing data (grey). As expected, the predictions conditional on mean estimates of fecundity (right) have narrow predictive intervals – they include only a subset of the uncertainty, i.e. that contributed by the seed data model assuming known fecundity. Predictions that incorporate the uncertainty in the state-space model itself (left) have broader predictive intervals and provide a more realistic prediction.

We constructed predictive intervals for a population and for individuals within the population, the latter conditioning on known covariates or estimated latent states associated with that individual. In principle, the predictive intervals obtained by methods discussed above should agree with the distributions not only of data (e.g. Figure 17.8), but also of latent states being estimated in the model. To illustrate that this is the case, Figure 17.9 provides an example that includes predictive intervals for growth, fecundity, and survival of *Quercus* where the latent states are represented as grey dots and predictive intervals are in black (a dark understory with $\lambda = 0.1$), grey (an intermediate exposure

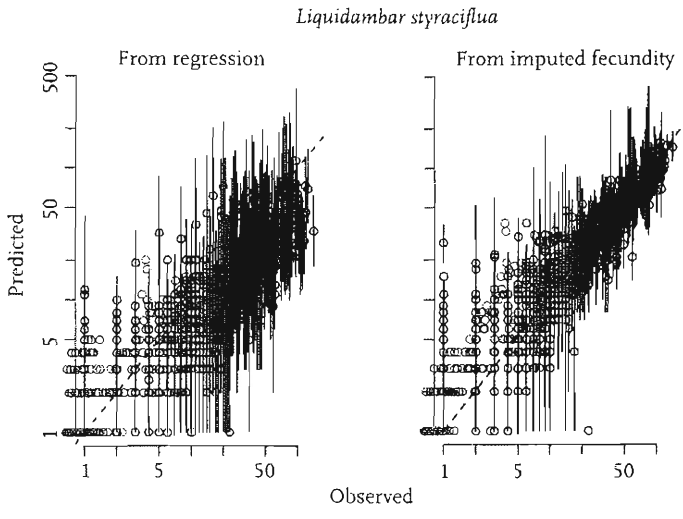

Fig. 17.8 Predictions for seed data conditioned on posterior mean estimates of fecundity (right) and on mean estimates of covariates (left). Predictive intervals are broader on the left, because they integrate not only uncertainty associated with dispersal and sampling, but also in the state-space model of fecundity.

level of $\lambda = 40$), and, for fecundity, dark grey (intermediate exposure, conditional on being mature). For black the sources of uncertainty, from the predictive mean outward are in order: parameter uncertainty (dashed – hardly visible), random individual effects (dotted – in this case small), year-to-year variation (dashed – in this case large), and process error (in this case small). In general we find agreement between estimates of latent states and the predicted variation from the model. The latent states for fecund individuals are covered by the predictive distributions conditional on being mature. The centre plot includes a large number of dots along the bottom of the plot, indicating immature individuals. The black, unconditional fecundity prediction marginalizes over the probability of being mature.

17.7 Summarizing the complexity

The large number of estimates generated by this analysis satisfies the need for a detailed representation of how demographic rates relate to one another, but it produces a new challenge: How do we summarize these results in meaningful ways? Despite the effort needed to simulate the distribution, obtaining the posterior is only the beginning. These results are now the basis for forward simulation experiments to determine how changing environments affect predictions of biodiversity. Here we simply point out the rich set of products that

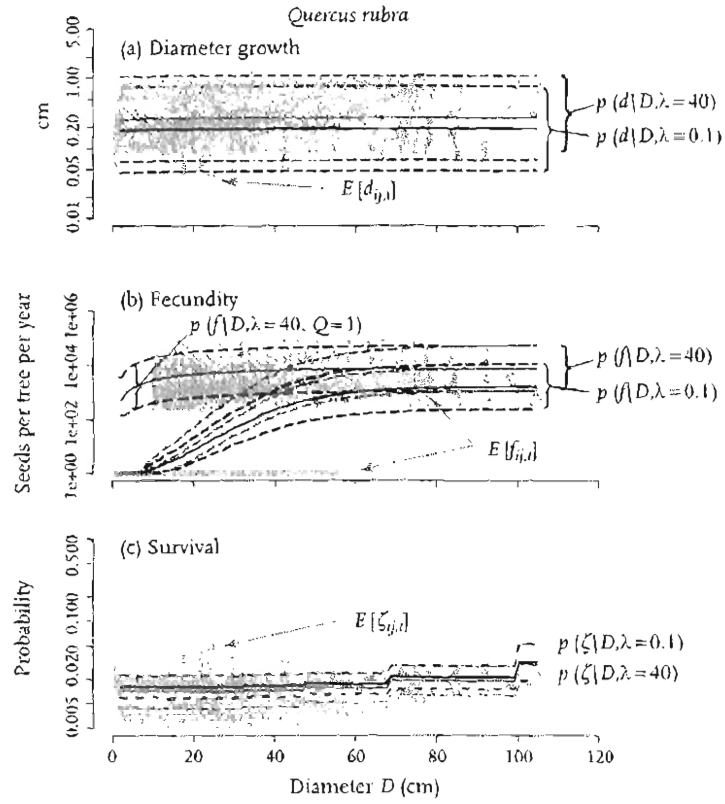

Fig. 17.9 Posterior mean estimates of latent states (dots) and predictive intervals for low ($\lambda = 0.1$; lower set of solid mean and dashed 95%) and high ($\lambda = 40$; upper set of solid mean and dashed 95%) canopy exposure. Included in (b) are predictive intervals for fecundity conditioned on mature status and high canopy exposure ($Q = 1$); zero values are jittered and plotted as one's to make them visible on this log scale.

can be derived from such results. For example, the predictive distributions provide a basis for simulation of interacting populations, a requirement for understanding how competition contributes to species diversity. The simulations in Figure 17.10 are conditioned on a particular assumption of light availability (in this case, a random draw from the data). Competition models generate the light availability based on shading from neighbours (e.g. Pacala *et al.* 1996, Govindarajan *et al.* 2004).

Despite the complexity of this analysis and the large numbers of estimates, predictions can be simple and valuable. For example, the decline in predictive intervals with increasing elevation for fecundity and growth rates in Figure 17.11 can help to explain mechanisms behind species range limits. Correlations among series of rates predicted from the model can be used to identify

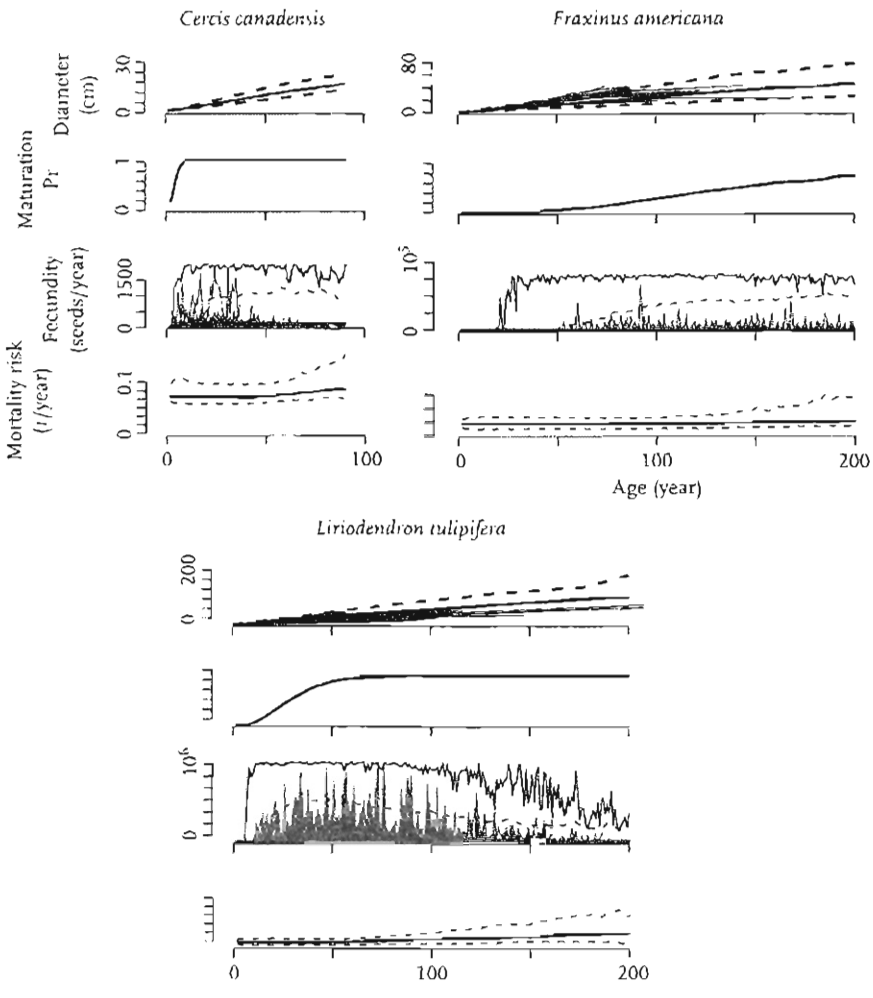

Fig. 17.10 Predictive simulations of demographic rates based on naïve scenarios for light availability, showing inherent differences among species. In each case are shown 95% predictive intervals, which include all sources of uncertainty in the model. For fecundity 30 individual simulations show the range of variability. Note that fecundity has a different scale for each species. For diameter, data are also shown (thin lines).

how lags in growth may affect fecundity (Figure 17.12, left side) and how and when rates of growth and fecundity deteriorate prior to death (Figure 17.12, centre and left). Predictive intervals for specific combinations of demographic rates (Figure 17.13) can be used to test hypotheses about the types of tradeoffs that might be needed for species to coexist. In each of these cases a seemingly incomprehensive number of estimates has been synthesized in ways that allow clear consideration of basic ecological relationships.

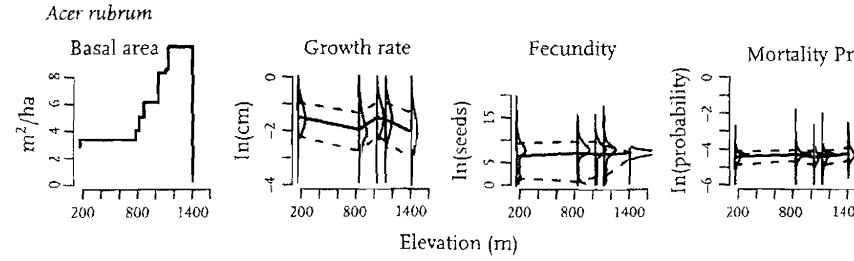

Fig. 17.11 Predictive intervals for demographic rates at different elevations. The predictive densities for different elevations (oriented horizontally) integrate to 1, with solid and dashed lines connecting posterior median and 95% CIs, respectively. The densities include parameter error and individual effects, but they do not include observation error.

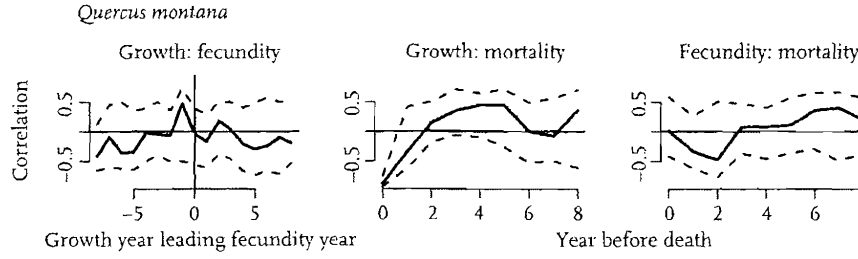

Fig. 17.12 Cross-correlations for demographic rates. The sequences of demographic rates for each individual were detrended and cross-correlated. Shown are bounds for 95% of the individuals (dashed) and medians (solid). In other words, dashed lines bound 95% of the individual cross-correlations.

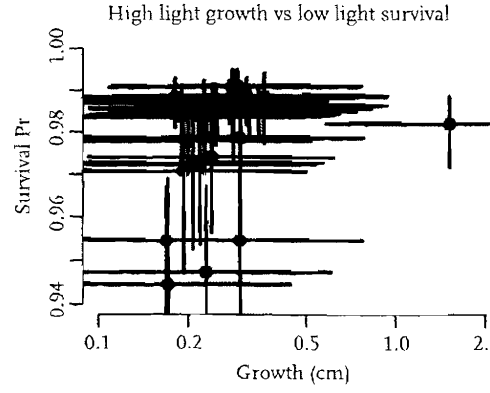

Fig. 17.13 Relationship among species in terms of capacity to grow fast at high light versus survival probability at low light. Predictive intervals are 95% and include only parameter error and variation among individuals.

17.8 Potential

Hierarchical modelling provides a mechanism for synthesis of complex information and interactions. The Bayesian framework is important for incorporation of prior knowledge, the strength of which differs for all parts of the model, including data and theoretical understanding. The advantage it provides for admitting the complexity unavoidable in the real world brings with it the challenge of understanding complex models. Once a large posterior is in hand, predictive distributions of key relationships can help to elucidate patterns of special interest.

Appendix

A. Broader context and background

A.1 Demographic models

Demographic modelling of natural populations has progressed substantially in recent years. Until a decade ago, models for inference on species in the wild involved a single stage, something like response = f(known inputs, stochastic error). The response could be growth rate or fecundity, the inputs being aspects of the individual's state, resources, or other factors that affect health. In fact, the most widely applied methods for inference involved a single predictor, the individual's stage in life, and it assumed that the response was linear. The approach owed its popularity to limited availability of covariate data and to readily available software. For a single response variable, such as fecundity f_i of tree i, linear models dominated. One might apply an allometric function of tree diameter D_i,

$$\begin{aligned} \ln f_i &= \alpha_0 + \alpha_0 \ln D_i + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2). \end{aligned}$$

Likewise, tree growth d_i might be modelled as a function of covariates x_i, such as light availability,

$$\begin{aligned} d_i &= f(x_i) + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2). \end{aligned}$$

As maximum likelihood has increased in popularity, nonlinear models have been more widely applied. For example, a growth rate might saturate with increasing resource availability. Ecologists have increasing developed code to perform optimizations (e.g. to obtain MLEs) and to simulate distributions (e.g. using a Metropolis algorithm). Mortality risk might be modelled as a function of past annual diameter increment d_i, because growth rate provides some

indication of overall tree health,

$$z_i \sim Bernoulli(\zeta_i)$$
$$\zeta_i = f(d_i)$$

using a standard GLM or some alternative form (Kobe *et al.* 1995, Yao *et al.* 2001). Such applications have been tremendously valuable and predispose ecologists to more advanced techniques, such as hierarchical modelling (Clark 2005, Latimer *et al.* 2006, Carlin *et al.* 2007).

The hierarchical application described in this chapter provides examples of many distributional forms and opportunity to mention some background for the hierarchical context. Hierarchical models are now being applied in ways that synthesize complex data to better understand species distributions (Latimer *et al.* 2006), migration (Wikle 2003, Hooten *et al.* 2003), mortality (Metcalf *et al.*, in review), disease spread (LaDeau *et al.* 2007) and environmental variation (Ibanez *et al.* 2007, Ogle *et al.* 2006, Dietze *et al.* 2008), growth (Clark *et al.* 2003, Mohan *et al.* 2007, Ibanez *et al.* 2007), and fecundity and dispersal (Clark *et al.* 2004, HilleRisLambers *et al.* 2006). Our chapter involves a specific application, emphasizing techniques that could be adapted for multiple applications, in part because it models demographic rates together.

A.2 GLMs in a hierarchical setting

Components of our model involve generalized linear models (GLMs) for maturation (Section 17.3.1) and seed data. GLMs involve an underlying linear predictor that is linked to a data distribution, typically binomial or Poisson, by a function that translates the predictor from the real line $(-\infty, \infty)$ to (0,1) for a binomial or $(0, \infty)$ for Poisson by way of a link function. As an example, the Bernoulli example in equation (17.1) shows $\theta_{ij,t}$ as an inverse logit function of the linear predictor $\beta_0^\theta + \beta_1^\theta D_{ij,t} + \beta_2^\theta \lambda_{ij,t}$. A large literature on GLMs dates at least to Nelder and Wedderburn (1972) with applications described for classical and Bayesian settings in many recent texts, including Gelman and Hill (2007). In the Bayesian context, parameters for fixed effects might have Gaussian priors. Our truncated normal priors (equation 17.21) do not importantly complicate the computation.

Hierarchically structured GLMs (or Generalized Linear Mixed Models) typically involve a stochastic specification of one or more terms in the linear predictor. Ecological applications often include random effects associated with location. For example, Latimer *et al.* (2006) include fixed effects as predictors of species occurrence at a geographic location, with a spatial random effect that introduces dependence based on proximity. The prior on this random effect depends on the value for neighbouring cells. This dependence introduces spatial smoothing. Ibanez *et al.* (2008) use random effects for location to absorb

some of the differences among stands not accounted for by covariates, in an application where stands are distant from one another. In both of these cases, the variance for random effects requires a prior, introducing an additional stage to the model.

Our application to the binary maturation process in equation (17.1) is applied to a time series for each individual tree, conditionally dependent on previous and future states. However, the underlying predictor is linear, with a logit link. Our approach differs from previous ones in placing the stochasticity not on the coefficients, but rather on the covariates. This structure (parameters as fixed effects assigned to random predictors) was motivated by the fact that tree size is estimated, and canopy exposure estimates are especially crude. The problem is constrained by informative priors on the predictor variables, particularly knowledge that they are bounded within a known range. This range enters as a truncated normal distribution (equations 17.12 and 17.19). Thus, the added stage involves sampling the covariates, rather than the coefficients.

The Poisson distribution for seed density (equation 17.4) differs from a standard GLM in that the expectation is a transport model for seed dispersal. More typically, applications would involve a linear predictor with a log link. Like the binomial, it may have a hierarchical specification. The inclusion of trap area as a coefficient in the Poisson piece of equation (17.3) is standard, accounting for the fact that collecting areas of traps differ.

Related to both the binomial and Poisson is the Zero-Inflated Poisson (*ZIPo*) model, a binomial-Poisson mixture, where the binomial piece can be interpreted as the probability that the Poisson-generating process exists (Lambert 1992, Hall 2000). For example, our fecundity model can be simplified to a simple probability of being mature given tree size (binary) and the expected fecundity, given that the individual is mature (LaDeau and Clark 2001),

$$p(f\,|\theta,\gamma) = [1-\theta+\theta Po(0\,|\gamma)]^{1(f=0)}\,[\theta Po(f\,|\gamma)]^{1(f>0)}$$

where f is seed production by a tree, $1()$ is the indicator function, θ is the probability of being mature, and γ is the fecundity given that an individual is mature. Marginally we expect $E(f) = \theta\gamma$. Note that the binary part, assigned probability θ, and the conditional Poisson fecundity with mean f are interpreted as different processes. We might include different predictors for θ and f or not (e.g. both depend on tree size and resources). This is a standard interpretation of a *ZIPo* model (e.g. Hall 2000, LaDeau and Clark 2001).

B. Models and computations

Gibbs sampling (Gelfand and Smith 1990) is widely used to simulate posteriors and is described in a number of recent texts (e.g. Carlin and Louis 2000, Gelman *et al.* 1995). In brief, one factors a high-dimensional posterior into a collection of

lower-dimensional densities that can be sampled successively, each conditional on others already updated. Conditionals may be sampled directly or indirectly using, for example, a Metropolis step. Here we describe conditional posteriors used for our Gibbs sampling algorithm. The general idea was to embed within the Gibbs sampler Metropolis steps, where direct sampling was not an option, with attention to blocking for efficiency.

B.1 State-space model

For the state space model (equation 17.7), all sampling was direct. The conditional posterior for fixed effect parameters is

$$vec(A)|\, X, Y, \Sigma, \ldots \sim N_{2p}(Vv, (\Sigma + V_b) \otimes V)\, I(a_1 < vec(A) < a_2)$$

where $V^{-1} = X^T X$, $v = X^T Z$, X is the stacked matrix of $(T_{ij} - t_{ij})$ by p X_{ij} matrices, one for each tree, $Z = [Z_{11}\ Z_{21} \ldots]$ is the stacked matrix $Z_{ij} = Y_{ij} - 1_{ij} b_t$, where 1_{ij} is the length $(T_{ij} - t_{ij})$ vector of 1's, and b_t are taken for the appropriate years in X and Z. The truncated multivariate normal is sampled from the conditional univariate truncated normals.

The error covariance matrix was sampled from an inverse Wishart conditional posterior. Let V_Σ be the parameter matrix for this prior. Then the conditional posterior for Σ^{-1} is

$$\Sigma^{-1} \sim W\left(\left[\sum_{i,j,t} Q_{ij,t}\left(y_{ij,t+1} - x_{ij,t}A - b_t - b_{ij}\right)'\left(y_{ij,t+1} - x_{ij,t}A - b_t - b_{ij}\right) + n_{IJT} V_\Sigma\right]^{-1}, \sum_{i,j,t} Q_{ij,t} + n_{IJT}\right)$$

where n_{IJT} is the total number of tree years (equation 17.16 of the text). The conditional posterior for the random effects variance is

$$\mathbf{V}_b^{-1} \sim W\left(\left[\sum_{i,j} \max_t\left(Q_{ij,t}\right) b_{ij}' b_{ij} + r_b R_b\right]^{-1}, \sum_{i,j} \max_t\left(Q_{ij,t}\right) + r_b\right).$$

Note that individual ij contributes to the conditional posterior only if it is imputed to be mature at some point during the study, in which case $\max_t(Q_{ij,t}) = 1$. The random effects are sampled from

$$b_{ij} \sim N_2(Vv, V),$$

where

$$V^{-1} = \left(T_{ij} - t_{ij}\right)\Sigma^{-1} + V_b^{-1},\ v = \Sigma^{-1} \sum_{t=\tau_{ij}}^{T_{ij}} \left(y_{ij,t+1}' - A' x_{ij,t}' - b_t\right),$$

τ_{ij} and T_{ij} are the first and last years during the study in which individual ij is imputed to be mature, and $T_{ij} - \tau_{ij}$ is the number of years mature that ij is mature during the study.

The fixed year effects are sampled from a conditional normal. Let V_t be the covariance matrix for the prior. The conditional posterior is $b_t \sim N_2(Vv, V)$, where $V^{-1} = \Sigma^{-1} \sum_{i,j} Q_{ij,t} + V_t^{-1}$, and $v = \Sigma^{-1} \sum_{i,j} Q_{ij,t}(y_{ij,t+1} - x_{ij,t} A - b_{ij})$. This draw was followed by subtraction of the mean for year effect for both $\ln d$ and $\ln f$. Of course b_t and A could be sampled with a single draw from a multivariate normal. A separate step was used due to the large number of b_t and the fact that each tree could have a different subset of total years.

B.2 Diameter growth

Diameter growth increments were updated from the conditional posterior given in equation (17.12) using a Gaussian approximation for the third factor, i.e. that corresponding to survival probability. That probability is approximated as $\zeta_{ij,t} \approx 1 - \mu_{d_{ij,t}}$, where the $\mu_{d_{ij,t}}$ sequence contains probabilities for discrete diameter increment bins. The contribution of diameter is omitted, because its contribution to survival probability is small relative to that of growth rate. Then the conditional distribution for the k bins is

$$p(\ln d_k | z) = \frac{p(z|\ln d_k)\, p(\ln d_k)}{\sum_k p(z|\ln d_k)\, p(\ln d_k)}$$

where $p(z = 1|\ln d_k) = 1 - \mu_k$ and $p(z = 0|\ln d_k) = \mu_k$, z being survival (1) or death (0) in the subsequent year, and $p(\ln d_k)$ the distribution of log growth rates. The conditional expectations and variances are

$$\mu_{d|z_{ij,t+1}} \equiv E(\ln d | z_{ij,t+1}) = \sum_k \ln d_k\, p(\ln d_k | z_{ij,t+1})$$

and

$$V_{d|z_{ij,t+1}} \equiv Var(\ln d | z_{ij,t+1}) = \sum_k (\ln d_k)^2\, p(\ln d_k | z_{ij,t+1}) - \left[E(\ln d | z_{ij,t+1})\right]^2.$$

There is a conditional mean and variance for $z = 0$ and $z = 1$. The log growth rates are sampled from $\ln d_{ij,t} \sim N(Vv, V)$, where

$$V^{-1} = V_{d|f}^{-1} + v_{ij,t}^{-1} + V_{d|z_{ij,t+1}}^{-1}, \quad v = \mu_{d|f} V_{d|f}^{-1} + \ln\left(d_{ij,t}^{(0)}\right) v_{ij,t}^{-1} + \mu_{d|z_{ij,t+1}} V_{d|z_{ij,t+1}}^{-1},$$

with conditional means and variances contributed by survival are for $z = 0$ or $z = 1$, depending on whether or not the individual survived until the next year.

B.3 Canopy exposure

Canopy values are sampled from a conditional posterior that depends on the prior means and variances coming from the analysis that assimilates data and

from the state-space model. If the individual is mature in year t, then canopy area is sampled from

$$\ln \lambda_{ij.t} \mid \sim N(V v, V)\, I\left(\left(c_{ij.t}^{(0)} - \sqrt{C_{ij.t}}\right) < \ln(\lambda_{ij.t}) < \left(c_{ij.t}^{(0)} + \sqrt{C_{ij.t}}\right)\right)$$

where $V^{-1} = A'_{\lambda\bullet} \Sigma^{-1} A_{\lambda\bullet} + C_{ij.t}^{-1}$, and $v = A_{\lambda\bullet}^{T} \Sigma^{-1} (y_{ij.t} - x_{ij.t}^{(-\lambda)} A_{(-\lambda\bullet)} - b_t - b_{ij}) + c_{ij.t}^{(0)} / C_{ij.t}$. Notation for the first factor follows that from the previous section. If immature in year t, $\ln \lambda_{ij.t}$ is sampled from a normal distribution having

$$V^{-1} = a_{\lambda}^{2} / w + C_{ij.t}^{-1}$$

$$v = \frac{\left(\ln(d_{ij.t}) - x_{ij.t}^{(-\lambda)} a_{(-\lambda)} - b_t - b_{ij}\right) a_{\lambda}}{w} + \frac{c_{ij.t}^{(0)}}{C_{ij.t}}$$

where $a_{(-\lambda)}$ is the vector a from equation (17.13), but lacking the coefficient for $\ln \lambda$, a_{λ} is the coefficient for $\ln \lambda$. w is the currently imputed error variance for growth rate regression, having an IG prior.

B.4 Fecundity, maturation, gender

Due to their conditional dependence structure, fecundity, maturation, and (for dioecious species) gender are sampled together in a Metropolis step. Here we describe sampling. The basic factoring used for maturation, gender, and fecundity is

$$p(f_{ij.t}, Q_{ij.t}, H_{ij}, s_{j.t} \mid q_{ij}, h_{ij}, d_{ij.t-1}, D_{ij.t}, \lambda_{ij.t}) = p(s_{j.t} \mid f_{ij.t}, Q_{ij.t}, H_{ij}) \times p(f_{ij.t}, Q_{ij.t}, H_{ij} \mid q_{ij}, h_{ij}, d_{ij.t-1} D_{ij.t}, \lambda_{ij.t})$$

where q_{ij} and h_{ij} represent the history of observations on individual ij, both past (before t) and future (after t). The first distribution on the right-hand side is the likelihood for seed trap data, indicating that all seed traps on plot j in year t conditionally depend on every tree i on plot j. The second factor on the right-hand side is the probability of being mature ($Q = 1$), female ($H = 1$), and having fecundity f.

For monoecious species, we use a Metropolis step where maturation status and fecundity are jointly proposed and rejected for all trees in a given stand j in a given year t. For dioecious species we must further sample gender. Because gender applies to an individual across all years, dioecious species are sampled in a different way and are discussed after monoecious species. The blocking differs between these two data types, which we describe here.

Efficient Gibbs sampling requires blocking of variables to facilitate mixing, which is challenging given the ways in which latent variables are linked with the unknown year in which an individual becomes mature τ_{ij}. These linkages include:

(i) the $Q_{ij,t}$ and $f_{ij,t}$ are inherently linked, by virtue of the fact that non-zero fecundity is defined only for mature individuals,

$$f_{ij,t}\,|(Q_{ij,t} = 0) = 0$$
$$f_{ij,t}\,|(Q_{ij,t} = 1) > 0;$$

(ii) maturation statuses for an individual over time are mutually dependent according to the one-way transition to maturity in year τ_{ij};
(iii) gender is considered to be fixed; and
(iv) seed trap data conditionally depend on all trees in the plot in a given year.

In light of the conditional relationships involving status and seed production, the choices for blocking are to (1) sample individually every year for every tree (conditioned on other trees for that year and other years for that individual), (2) sample as a block all individuals within a plot for a given year, and (3) sample as a block all trees and years within a plot. The first option has the advantage that high acceptance rates can be achieved, but is computationally slow, entailing loops over plots, individuals, and years, e.g. a Metropolis step for every tree-year in the data set. The third option necessarily results in a high rejection rate, each proposal consisting of $\sum_{i=1}^{n_j}(T_{ij} - t_{ij})$ values. The binary nature of Q and H proposals can make acceptance rates low. Nonetheless, because gender H_{ij} applies to an individual across all years, we use a modification of option 3 for dioecious species. We begin with a description for monoecious species, followed by the description for dioecious species.

Monoecious species – We use the second option for monoecious species, blocking on time and modelling each year successively. The factoring is

$$p(f_{j,t}, Q_{j,t}\,|q_{j,t}, x_{j,t-1}, x_{j,t}, Q_{j,t-1}, Q_{j,t+1}, s_{j,t}) \propto p(s_{j,t}\,|f_{j,t}, Q_{j,t})$$
$$p(Q_{j,t}, f_{j,t}\,|q_{j,t}, x_{j,t-1}, Q_{j,t-1}, Q_{j,t+1}, x_{j,t}).$$

We propose all values of $\{Q, f\}_{j,t}$ together and accept or reject them as a block. The Markov transition probabilities from t to $t+1$ are conditioned on observations of status, and they must be combined with probabilities for fecundities and seed trap data. The transition from immature to mature is a hidden Markov process, but only for tree-years in which status is unknown, which is the case after the last year in which immaturity is certain, and before reproduction has been observed. If the status is known through past observations (if previously observed to be mature, then still mature), a current observation (mature or immature), or future observations (if later known to be immature, then immature now), then status $Q_{ij,t}$ is known. This is also the case for imputed statuses. If unknown, status must be modeled as the conditional probability of being in state $Q_{ij,t} = 0$ or 1 given $Q_{ij,t-1}$ and $Q_{ij,t+1}$. These probabilities involve the age-specific rates of making the transition from immature to mature states and can

be derived from the cumulative logit probability of being mature given diameter $D_{ij,t}$ and canopy status $\lambda_{ij,t}$ (equation 17.2a). Because blocking is year-by-year, we condition the transition probability on both the foregoing and the following years. Then the trivial probabilities are

$$p\left(Q_{ij,t} = 1 \mid Q_{ij,t-1}\right) = 1$$
$$p\left(Q_{ij,t} = 1 \mid Q_{ij,t+} = 0\right) = 0.$$

For failure to recognize the reproductive state, we need the additional factor

$$p\left(q_{ij,t} = 0 \mid Q_{ij,t} = 1\right) = 1 - \upsilon.$$

For Gibbs sampling, we need the year-by-year transition probabilities from immature to mature between $t-1$ and t given that the transition was made between $t-1$ and $t+1$. Let $\delta_{ij,t}$ be the probability of being in the mature state conditional on states in years $t-1$, $t+1$, and on observations. Ignoring observations for the moment, we have

$$\begin{aligned}
\delta_{ij,t} &= p\left(Q_{ij,t} = 1 \mid Q_{ij,t-1} = 0,\ Q_{ij,t+1} = 1\right) \\
&= \frac{p\left(Q_{ij,t} = 1 \mid Q_{ij,t-1} = 0\right) p\left(Q_{ij,t+1} = 1 \mid Q_{ij,t} = 1\right)}{\sum_{k=0,1} p\left(Q_{ij,t} = k \mid Q_{ij,t-1} = 0\right) p\left(Q_{ij,t+1} = 1 \mid Q_{ij,t} = k\right)} \\
&= \frac{d\theta_{ij,t}/\left(1-\theta_{ij,t-1}\right)}{d\theta_{ij,t}/\left(1-\theta_{ij,t-1}\right) + \left[1 - d\theta_{ij,t}/\left(1-\theta_{ij,t-1}\right)\right] \times d\theta_{ij,t+1}/\left(1-\theta_{ij,t}\right)} \\
&= \frac{d\theta_{ij,t}}{d\theta_{ij,t} + \left(\frac{1-\theta_{ij,t-1}-d\theta_{ij,t}}{1-\theta_{ij,t}}\right) d\theta_{ij,t+1}} \\
&= \frac{d\theta_{ij,t}}{d\theta_{ij,t} + d\theta_{ij,t+1}}
\end{aligned}$$

where

$$\begin{aligned}
d\theta_{ij,t} &= \left(\frac{d\theta_{ij,t}}{d\,D_{ij,t}} \times \frac{d\,D_{ij,t}}{dt}\right) dt = \left(\frac{d\theta_{ij,t}}{dD_{ij,t}} \times d_{ij,t}\right) dt \\
&= \beta_1^{\theta} d_{ij,t} \theta_{ij,t}\left(1-\theta_{ij,t}\right) dt.
\end{aligned}$$

Because $\lambda_{ij,t}$ changes much slower than $D_{ij,t}$, we do not include it in the chain rule calculation for the derivative. Because dt is always equal to 1 year, we hereafter omit it.

Observations change the transition probabilities. The previous equation for $d_{ij,t}$ describes the probability of transition in the absence of an observation. If there is an observation in year t and it is 'uncertain' ($q_{ij,t} = 0$: see Table 17.4), then the observer did not identify the tree as mature, and the probability

becomes

$$\delta_{ij,t} = \Pr\left(Q_{ij,t} = 1 \,\middle|\, Q_{ij,t-1} = 0,\ Q_{ij,t+1} = 1,\ q_{ij,t} = 0\right) = \frac{d\theta_{ij,t}(1-\upsilon)}{d\theta_{ij,t}(1-\upsilon) + d\theta_{ij,t+1}}.$$

For the first study year, in the absence of an observation (maturation statuses were not obtained on all individuals the first year of the study), we have

$$\delta_{ij,t} = p\left(Q_{ij,t} = 1 \,\middle|\, Q_{ij,t+1} = 1\right) = \frac{\theta_{ij,t}}{\theta_{ij,t} + d\theta_{ij,t+1}}.$$

If there was an observation and that observation was $q_{ij,t} = 0$ (Table 17.4), this becomes

$$\frac{\theta_{ij,t}(1-\upsilon)}{\theta_{ij,t}(1-\upsilon) + d\theta_{ij,t+1}}.$$

For the last observation year, absent observation,

$$\delta_{ij,T} = p\left(Q_{ij,T} = 1 \,\middle|\, Q_{ij,T-1} = 0\right) = \frac{d\theta_{ij,T}}{\left(1-\theta_{ij,T}\right)}.$$

If there was an observation, we have

$$\delta_{ij,T} = \frac{d\theta_{ij,T}(1-\upsilon)}{\left(1-\theta_{ij,T}\right)}.$$

The Metropolis steps entail a loop over time (17 year), at each time step proposing values for Q_t and f_t, with the constraints on Q discussed above and $f_{ij,t} = 0$ for all imputed $Q_{ij,t} = 0$. For those imputed to be immature at $t-1$, candidate values come from $Q^*_{ij,t} \sim Bernoulli(0.5)$. All others remain mature. For individuals previously imputed to be mature, we propose $\ln f^*_{ij,t} \sim N(\ln f^{(g)}_{ij,t}, 0.1)$, where (g) denotes the current Gibbs step, prior to updating. For individuals previously imputed to be immature, but now mature, we propose from $f^*_{ij,t} | (Q_{ij,t-1} = 0) \sim N(\ln f'_{ij,t}, 0.1)$, where $f'_{ij,t}$ is an auxiliary variable having the value retained from the most recent iteration of the Gibbs sampler in which $Q_{ij,t} = 1$. The acceptance criterion involves the products from equation (17.2a),

$$\begin{aligned} &Q_{j,t},\ f_{j,t} \,\middle|\, Q_{j,t-1},\ Q_{j,t+1},\ q_{j,t},\ s_{j,t},\ \mu^{f|d}_{j,t},\ V^{f|d}_{j,t} \\ &\quad \propto \prod_i \left(1-\delta_{ij,t}\right)^{1-Q_{ij,t}} \left[\delta_{ij,t} N\left(\ln\left(f_{ij,t}\right) \,\middle|\, \mu^{f|d}_{ij,t},\ V^{f|d}_{ij,t}\right)\right]^{Q_{ij,t}} \\ &\quad \times \prod_{k=1} Po\left(s_{jk,t} \,\middle|\, A_{jk} g_{jk}\left(f_{j,t}; u\right)\right). \end{aligned}$$

Note that all individuals imputed to be mature have a conditional density associated with the state-space model. The $\delta_{ij,t}$ are different for each individual and year, as discussed above. All trees contribute to the likelihood for the seed data for plot j in year t, in that by producing seed or not, they affect the parameters of the Poisson sampling distribution for seed. Of course, the set of proposals is accepted or rejected as a block. The sampler is more efficient than it appears, because we can propose statuses and fecundities for all plots simultaneously, and accept/reject them on a plot by plot basis, without actually looping over plots. Once states are updated for time t, we move to $t + 1$.

Dioecious species – For dioecious species gender is unchanging over time, so we evaluate the full history of observations for each tree, but still avoiding loops over individual trees. It is efficient to factor the conditional somewhat differently, taking together all trees on plot j over all years,

$$p(f_j, Q_j, H_j \,|q_j, h_j, X_j, s_j) \propto p(s_j \,|f_j, Q_j, H_j)\, p(f_j, Q_j, H_j \,|q_j, h_j, X_j).$$

Because maturation is no longer modeled year-by-year, we require the probability for a history of maturation status, conditioned on observations obtained sporadically over individuals and years.

Let τ_{ij} be the year in which an individual becomes mature, $\tau_{ij}^0 = \max_t(q_{ij,t} = -1)$ be the last year an individual is known to have been immature, and $\tau_{ij}^1 = \min_t(q_{ij,t} = 1)$ be the first year an individual is known to have been mature. Thus, we have the constraint $\tau_{ij}^0 \le \tau_{ij} \le \tau_{ij}^1$. The probability assigned to an individual that became mature in year t is,

$$\begin{aligned}\delta_{ij} &= p\left(\tau_{ij} = t \,\middle|\, \tau_{ij}^0 < \tau_{ij} < \tau_{ij}^1\right) = d\theta_{ij,t} \Big/ \left(\theta_{ij,\tau_{ij}^1} - \theta_{ij,\tau_{ij}^0}\right) \\ &= \beta_1^\theta d_{ij,t} \frac{\theta_{ij,t}(1 - \theta_{ij,t})}{\left(\theta_{ij,\tau_{ij}^1} - \theta_{ij,\tau_{ij}^0}\right)}.\end{aligned}$$

For individuals imputed to be still immature at the end of the observation period at T_{ij}, the probability is

$$\delta_{ij} = p\left(\tau_{ij} > T_{ij} \,\middle|\, \tau_{ij} > \tau_{ij}^0\right) = 1 - p\left(\tau_{ij} \le T_{ij} \,\middle|\, \tau_{ij} > \tau_{ij}^0\right) = \frac{1 - \theta_{T_{ij}}}{1 - \theta_{ij,\tau_{ij}^0}}.$$

For individuals imputed to be already mature before observations began at t_{ij}, the probability is

$$\delta_{ij} = p\left(\tau_{ij} < t_{ij} \,\middle|\, \tau_{ij} < \tau_{ij}^1\right) = \frac{\theta_{t_{ij}}}{\theta_{ij,\tau_{ij}^1}}.$$

We now have a probability for the history of an individual that became mature at time τ_{ij} or remained immature throughout. Our prior specification allows for

a minimum and maximum maturation diameter, in which case θ_{ij,τ_{ij}^0} and θ_{ij,τ_{ij}^1} are the values of θ taken at these prior maturation diameter values.

The probability for maturation is combined with observations of status between those years that established it as immature and mature and for gender. Thus far, we have considered observations that definitively establish maturity or immaturity ($q_{ij,t} = -1$ or 1 in Table 17.4). For $q_{ij,t} = 0$, status is uncertain. Status detection is defined as $p(q_{ij,t} = 1 | Q_{ij,t} = 0) = 0$ and $p(q_{ij,t} = 1 | Q_{ij,t} = 1) = v$. The individual has unknown gender if the gender is not observed, the observation is uncertain, or if flowers are observed but not identified to sex, and no observation is available from the fruiting season (Table 17.4). Considering both observations and gender, the probability for individual ij becomes

$$p\left(\tau_{ij}, H_{ij} \middle| \tau_{ij}^0 < \tau_{ij} < \tau_{ij}^1, q_{ij}\right) = \delta_{ij}(1-v)^{n_{ij}^v}(1-\phi)^{1-H_{ij}}\phi^{H_{ij}}$$

where n_{ij}^v is defined to be the number of times mature status was 'undetected' during the interval (τ_{ij}, τ_{ij}^1), i.e. the number of times that an individual imputed to be mature in year τ_{ij} was not identified as such. If gender is known, the third factor disappears. The full reproductive history on all individuals has conditional probability

$$p\left(Q_j, H_j, f_j \middle| q_j, h_j, s_j, \mu_j^{f|d}, V_j^{f|d}\right) = \prod_i \delta_{ij}(1-v)^{n_{ij}^v}(1-\phi)^{1-H_{ij}}$$

$$\times \left[\phi \prod_t N\left(\ln(f_{ij,t}) \middle| \mu_{ij,t}^{f|d}, V_{ij,t}^{f|d}\right)^{Q_{ij,t}}\right]^{H_{ij}} \prod_t \prod_k Po\left(s_{jk,t} \middle| A_{jk} g_{jk}(f_{j,t}; u)\right).$$

For each individual a maturation diameter is proposed from a uniform distribution

$$\tau_{ij} \sim unif\left(t_{ij}^0, t_{ij}^1\right)$$

$$t_{ij}^0 = \max\left(t_{ij}^{\min}, \tau_{ij}^0\right)$$

$$t_{ij}^1 = \min\left(t_{ij}^{\max}, \tau_{ij}^1\right).$$

The bounds for minimum and maximum maturation diameters are not sooner than the first year in which ij reached the minimum prior diameter for maturation $t_{ij}^{\min} = \max(t | (D_{ij,t} > D_{\min}))$ or it was last known to be immature τ_{ij}^0 and not later than the last year in which ij had not yet reached the prior diameter for certain maturation $t_{ij}^{\max} = \min(t | (D_{ij,t} < D_{\max}))$ or it was known to be mature τ_{ij}^1. There are prior minimum and maximum diameters, which differ among species. The fecundity for an individual proposed to be immature is zero. For individuals currently imputed to be mature, the proposed fecundity is a truncated normal on $(f_{\min}, f_{\max})$ centered on the current estimate. The conditional

densities are then the product of Poisson seed data, Gaussian fecundity, and the probability associated with maturation in year t. Because the probability of seed data conditionally depends on all trees in the stand in all years, the ensemble of (f_j, Q_j, H_j) is accepted or rejected as a block.

Recognition error is sampled from

$$Bin\left(v^{(1)} \,|v^{(0)} + v^{(1)}, v\right) Be(v\,|v_1, v_2) = Be\left(v\,|v_1 + v^{(1)}, v_2 + v^{(0)}\right)$$

where the two arguments are sums of prior values and numbers of currently imputed mature individuals for which maturation was recognized as such or not, i.e.

$$v^{(0)} = \sum_{ij,t} I\left(q_{ij.t} = 0,\ Q_{ij.t} = 1\right)$$

$$v^{(1)} = \sum_{ij.t} I\left(q_{ij.t} = 1,\ Q_{ij.t} = 1\right).$$

For the female fraction we sample from the conditional posterior

$$Bin\left(h^{(1)} \,|h^{(0)} + h^{(1)}, \phi\right) Be(\phi\,|h_1, h_2) = Be\left(\phi\,|h_1 + h^{(1)}, h_2 + h^{(0)}\right)$$

where the two arguments are sums of prior and currently imputed numbers of the females and males, respectively,

$$h^{(0)} = \sum_{ij}\left(1 - H_{ij}\right)$$

$$h^{(1)} = \sum_{ij} H_{ij}.$$

Prior values are $h_1 = h_2 = 4$, which has a mean of 0.5 and is dominated by the data.

Parameters for the logit function of maturation equation (17.1) are sampled with Metropolis step. Conditionally we have

$$\prod_{i,j} \delta\left(\tau_{ij}\right) N\left(\beta^{\theta}\,|b^{\theta}, V^{\theta}\right) I\left(\beta^{\theta}_{2,3} > 0\right)$$

where the first product is the probability associated with maturation years, which depend on β^{θ}, and the truncated normal prior. A proposal is generated from a multivariate normal truncated at zero for these two parameters.

Parameters for the seed data model are sampled in a single Metropolis step. Conditionally,

$$\prod_{t=t_j}^{T_j} \prod_{k=1}^{K_j} Po\left(s_{jk.t}\,|A_{jk} g_{jk}\left(f_{j.t}; u, c\right)\right) N(u\,|u_0, V_u)\ N(c\,|c_0, V_c)\ I(u, c > 0)\,.$$

Values are proposed from a normal distribution. In the case of missing data, seed counts were replaced with the currently imputed seed value.

Imputation of missing data involved a Metropolis step with proposals of plus or minus 1 from the current value with probability 0.5. The conditional posterior includes a Poisson prior with a mean density as discussed in Section 17.4 multiplied by Poisson density for sample *jk,t*. Proposals were accepted as a block for $s_{j,t}$.

A Metropolis step is used to simultaneously update all of the diameter growth and diameter bins for the nonparametric survival relationship. The growth rates and diameters of all individuals are binned in the sequences μ_d and μ_D are for the all years. For diameter increment there are 31 bins equally spaced with width 0.1 on the $\log_{10}$ scale. For diameter there are six bins, also equally spaced on the $\log_{10}$ scale, with the maximum value chosen to exceed that largest diameter in the data set. Survival from year t to $t+1$ is the event $z_{ij,t} = 1$ and death in the subsequent year is $z_{ij,t} = 0$. At each Gibbs step new sequences of μ_d and μ_D are proposed each being Gaussian and centred on the currently imputed values, but truncated midway between the current values. For diameter increment the proposal distribution is

$$N\left(\mu_d^* \,|\, \mu_d,\, V\right) I\left(\left(\mu_{d-1} - \mu_d\right)/2 < \mu_d^* < \left(\mu_d - \mu_{d+1}\right)/2\right)$$

where V is a small value (0.1 in this case). In other words, if the currently imputed value for $\mu_{d,k}$ was 0.5 and those for bins $k-1$ and $k+1$ were 0.6 and 0.48, then the proposal would come from the normal centred at 0.5, truncated at 0.55 and 0.49. This procedure allows for any shape subject to monotonic decline. The proposals for the diameter values are done in the same way, with the constraint being monotonic increase and with an informative prior $Be(a_k, b_k)$. All values (both growth rate and diameter) are proposed together and accepted as a block.

References

Batista, W. B., Platt, W. J. and Macchiavelli, R. E. (1998). Demography of a shade-tolerant tree (*Fagus grandifolia*) in a hurricane-disturbed forest. *Ecology*, **79**, 38–53.

Beckage, B. and Clark, J. S. (2003). Seedling survival and growth in southern Appalachian forests: Does spatial heterogeneity maintain species diversity? *Ecology*, **84**, 1849–1861.

Bigler, C. and Bugmann, H. (2004). Predicting the time of tree death using dendrochronological data. *Ecological Applications*, **14**, 902–914.

Carlin, B., Clark, J. S. and Gelfand, A. (2006). Elements of Bayesian Inference. In *Hierarchical Models of the Environment*, (ed. J.S. Clark and A. Gelfand), pp. 3–24. Oxford University Press, Oxford.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, Boca Raton, FL.

Caswell, H. (2001). *Matrix Population Models*. Sinauer, Sunderland, MA.

Clark, J. S., LaDeau, S. and Ibanez, I. (2004). Fecundity of trees and the colonization-competition hypothesis, *Ecological Monographs*, **74**, 415–442.

Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecology Letters*, **8**, 2–14.

Clark, D. B. and Clark, D. A. (1996). Abundance, growth and mortality of very large trees in neotropical lowland rain forest. *Forest Ecology and Management*, **80**, 235–244.

Clark, J. S., Macklin, E. and Wood, L. (1998). Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs*, **68**, 213–235.

Clark, J. S., Silman, M., Kern, R., Macklin, E. and Hille Ris Lambers, J. (1999). Seed dispersal near and far: Generalized patterns across temperate and tropical forests. *Ecology*, **80**, 1475–1494.

Clark, J. S., Wolosin, M., Dietze, M., Ibanez, I., LaDeau, S., Welsh, M. and Kloeppel, B. (2007). Tree growth inferences and prediction from diameter censuses and ring widths. *Ecological Applications*, **17**, 1942–1953.

Condit, R., Hubbell, S. P. and Foster, R. B. (1995). Mortality rates of 205 neotropical tree and shrub species and the impact of severe drought. *Ecological Monographs*, **65**, 419–439.

Coomes, D. A. and Allen, R. B. (2007). Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology*, **95**, 27–40.

Dietze, M., Wolosin, M. and Clark, J. (2008). Capturing diversity and individual variability in allometries: A hierarchical approach. *Forest Ecology and Management*, **256**, 1939–1948.

Govindarajan, S., Dietze, M., Agarwal, P. and Clark, J. S. (2004). A scalable model of forest dynamics. Proceedings of the 20th Symposium on Computational Geometry SCG, pp. 106–115.

Gurevitch, J., Scheiner, S. M. and Fox, G. A. (2002). *The Ecology of Plants*. Sinauer, Sunderland, MA.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Gelman, A. and Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

Gurevitch, J., Scheiner, S. M. and Fox, G. A. (2002). *The Ecology of Plants*. Sinauer, Sunderland, MA.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, **56**, 1030–1039.

Hooten, M. B., Larsen, D. R. and Wikle, C. K. (2003). Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, **18**, 487–502.

Ibáñez, I., Clark, J. S. and Dietze, M. (2007). Evaluating the sources of potential migrant species: Implications under climate change. *Ecological Applications*, **18**, 1664–1678.

Ibáñez, I., Clark, J. S., LaDeau, S. and Hille Ris Lambers, J. (2007). Exploiting temporal variability to understand tree recruitment response to climate change. *Ecology Monographs*, **77**, 163–177.

King, D. A., Davies, S. J. and Noor, N. S. M. (2006). Growth and mortality are related to adult tree size in a Malaysian mixed dipterocarp forest. *Forenerty and Ecological Management*, **223**, 152–158.

Kobe, R. K., Pacala, S. W., Jr. Silander, J. A. and Canham, C. D. (1995). Juvenile tree survivorship as a component of shade tolerance. *Ecological Applications*, **5**, 517–532.

Kobe, R. K. (2006). Sapling growth as a function of light and landscape-level variation in soil water and foliar N in northern Michigan. *Oecologia*, **147**, 119–133.

LaDeau, S. L. and Clark, J. S. (2006). Elevated CO_2 and tree fecundity: The role of tree size, interannual variability, and population heterogeneity. *Global Change Biology*, **12**, 822–833.

LaDeau, S., Kilpatrick, M. and Marra, P. P. (2007). West Nile virus emergence and large-scale declines of North American bird populations. *Nature*, **447**, 710–714.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1–14.

Latimer, A. M., S. Wu, A. E. Gelfand, and J. A. Silander. (2006). Building statistical models to analyze species distributions, *Ecological Applications*, **16**, 33–50.

Metcalf, C. J. E., Clark, J. S. and McMahon, S. M. (2009). Overcoming data sparseness and parametric constraints in modeling of tree mortality: A new non-parametric Bayesian model. *Canadian Journal of Forest Research*, **39**, 1677–1687.

Mohan, J. E., Clark, J. S. and Schlesinger, W. H. (2007). Long-term CO_2 enrichment of an intact forest ecosystem: Implications for temperate forest regeneration and succession. *Ecological Applications*, **17**, 1198–1212.

Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P. and Cardinot, G. (2007). Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology*, **88**, 2259–2269.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society* (Series A), **135**, 370–384.

Ogle, K., Uriarte, M., Thompson, J., Johnstone, J., Jones, A., Lin, Y., McIntire, E., Zimmerman, J. (2006). Implications of vulnerability to hurricane damage for long-term survival of tropical tree species: A Bayesian hierarchical analysis. In *Hierarchical Modeling for the Environmental Sciences: Statistical Methods and Applications*, (ed. J.S. Clark and A.E. Gelfand), Oxford University Press, Oxford.

Rich, R. L., Frelich, L. E. and Reich, P. B. (2007). Wind throw mortality in the southern boreal forest: Effects of species diameter and stand age. *Journal of Ecology*, **95**, 1261–1273.

Suarez, M. L., Ghermandi, L. and Kitzberger, T. (2004). Factors predisposing episodic drought-induced tree mortality in Nothofagus site, climatic sensitivity and growth trends. *Journal of Ecology*, **92**, 954–966.

Thomas, S. C. (1996). Relative size at the onset of maturity in rain forest trees: A comparative analysis of 37 Malaysian species. *Oikos*, **76**, 145–154.

Uriarte, M., Canham, C. D., Thompson, J. and Zimmerman, J. K. (2004). A neighborhood analysis of tree growth and survival in a hurricane-driven tropical forest. *Ecological Monographs*, **74**, 591–614.

Valle, D., Clark, J. S., Dietze, M., Agarwal, P. K., Millette, T. and Schultz, H. (2009). The effect of light competition on growth rate of large trees. In preparation.

van Mantgem, P.J. and Stephenson, N. L. (2007). Apparent climatically induced increase of tree mortality rates in a temperate forest. *Ecology Letters*, **10**, 909–916.

Wyckoff, P. H., and Clark, J. S. (2000). Predicting tree mortality from diameter growth: A comparison of maximum likelihood and Bayesian approaches *Canadian Journal of Forest Research*, **30**, 156–167.

Wikle, C. K. (2003). Hierarchical Bayesian models of predicting the spread of ecological processes. *Ecology*, **84**, 1382–1394.

Yao X., Titus, S. J. and MacDonald, S. E. (2001). A generalized logistic model of individual tree mortality for aspen, white spruce, and lodgepole pine in Alberta mixedwood forests. *Canadian Journal of Forestry Research*, **31**, 283–291.