Overcoming data sparseness and parametric constraints in modeling of tree mortality: a new nonparametric Bayesian model

C. Jessica E. Metcalf, Sean M. McMahon, and James S. Clark

Abstract: Accurately describing patterns of tree mortality is central to understanding forest dynamics and is important for both management and ecological inference. However, for many tree species, annual survival of most individuals is high, so that mortality is rare and, therefore, difficult to estimate. Furthermore, tree mortality models have potentially complex suites of covariates. Here, we extend traditional and recent approaches to modeling tree mortality and propose a new nonparametric Bayesian method. Our model is constrained to both reflect and distinguish known relationships between mortality and its two key covariates, diameter and diameter increment growth, but it remains sufficiently flexible to capture a wide variety of patterns of mortality across these covariates. Our model also allows incorporation of outside information in the form of priors, so that increased mortality of large trees can always be formally modeled even when data are sparse. We present results for our nonparametric Bayesian mortality model for maple (*Acer* spp.), holly (*Ilex* spp.), sweet gum (*Liquidambar styraciflua* L.), and tulip-poplar (*Liriodendron tulipifera* L.) populations from North Carolina, USA.

Résumé : La description exacte des patrons de mortalité des arbres est essentielle pour comprendre la dynamique forestière et importante pour les inférences écologiques et celles qui concernent l'aménagement. Cependant, chez plusieurs espèces d'arbre la survie annuelle de la plupart des individus est élevée, de telle sorte que la mortalité est rare et, par conséquent, difficile à estimer. De plus, les modèles de mortalité des arbres comportent des suites de covariables potentiellement complexes. Dans cet article, nous étendons la portée des approches récentes et traditionnelles pour modéliser la mortalité des arbres et nous proposons une nouvelle méthode bayésienne non paramétrique. Notre modèle est contraint tant pour refléter que pour distinguer les relations connues entre la mortalité et ses deux covariables clés : le diamètre et la croissance en diamètre, mais il demeure suffisamment flexible pour reconnaître une large variété de patrons de mortalité parmi toutes les valeurs de ces covariables. Notre modèle permet également d'incorporer de l'information externe sous forme d'a priori, de telle sorte que l'augmentation de la mortalité des gros arbres puisse toujours être formellement modélisée même lorsque les données sont peu nombreuses. Nous présentons les résultats de notre modèle bayésien non paramétrique de mortalité pour des populations d'érable (*Acer* spp.), de houx (*Ilex* spp.), de liquidambar (*Liquidambar styraciflua* L.) et de tulipier (*Liriodendron tulipifera* L.) de la Caroline du Nord, aux États-Unis d'Amérique.

[Traduit par la Rédaction]

Introduction

Tree mortality directly and indirectly determines population turnover (Condit et al. 1993; Clark and Clark 1999; Clark et al. 2007), species coexistence (Condit et al. 2006; Gilbert et al. 2006), carbon sequestration (Caspersen et al. 2000), forest size distributions (Coomes and Allen 2007), and forest responses to climate change (Clark 2002, 2004). Two commonly used predictors for modeling mortality are diameter growth increment and diameter. Individuals with small diameter growth increments generally also have a high probability of mortality (Condit et al. 1995; Kobe et al. 1995; Wyckoff and Clark 2002). Diameter growth increment is thought to integrate the multiple stresses faced by a tree, including drought (Suarez et al. 2004), fungal infections (Cherubini et al. 2002), insect pests (Burleigh et al. 2002), and low light levels (Condit et al. 1995; Pacala et al. 1996; Wyckoff and Clark 2002). The other commonly used covariate of mortality, diameter, is associated with high mortality at the extremes of diameter ranges, creating a U-shaped response to mortality (Platt et al. 1988). Small trees suffer high mortality chiefly because their small size renders them more vulnerable to increasing resource limitation through shading (Wyckoff and Clark 2002); thus small-tree mortality is largely associated with slow growth. Large trees have increased mortality rates (e.g., Monserud and Sterba 1999; King et al. 2006; Coomes and Allen 2007) for structural reasons whose mechanistic foundations are increasingly

Received 9 September 2008. Accepted 12 May 2009. Published on the NRC Research Press Web site at cjfr.nrc.ca on 9 September 2009.

C.J.E. Metcalf.[1,2] Centre for Infectious Disease Dynamics, Pennsylvania State University, State College, PA 16801, USA.
S.M. McMahon. Smithsonian Tropical Research Institute, Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater, MD 21307-0028, USA.
J.S. Clark. Smithsonian Tropical Research Institute, Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater, MD 21307-0028, USA; Department of Biology, Duke University, Durham, NC 27707, USA; Department of Statistical Science, Duke University Durham, NC 27707, USA.

[1]Corresponding author (e-mail: cjm29@duke.edu).
[2]Present address: : Department of Ecology and Evolutionary Biology and Woodrow Wilson School, Eno Hall, Princeton University, Princeton, NJ 08544-2016, USA.

well described (Yoda et al. 1965; Ryan and Yoder 1997; Koch et al. 2004). Capturing this U-shaped pattern of mortality over diameter is key to modeling tree mortality not only because it is an accurate reflection of the biology but also because the goal of many studies of tree mortality is to obtain parameters for forest simulators (e.g., Monserud and Sterba 1999; Eid and Tuhus 2001). As the death of large canopy trees is a rare event relative to the temporal and spatial scale of research studies, the full U-shaped pattern is often difficult to detect (Monserud and Sterba 1999; Nord-Larsen 2006).

Researchers have attempted to capture the complex shape of mortality across diameter at breast height, D, by using a range of different polynomial and hyperbolic forms as covariates within a logistic regression. For example, in a highly cited paper, Monserud and Sterba (1999) use D^{-1}, D, and D^2 to model mortality of trees in Austrian forests. Recent examples of logistic regression with various forms of diameter as a covariate include Wunder et al. (2007), Zhao et al. (2004), and Rich et al. (2007). However, logistic regressions are sensitive to the distribution of data (Lavine 1991; Wyckoff and Clark 2002); hence, for data sets with restricted distributions of tree diameter, extrapolation of the pattern of mortality across diameter may be problematic (e.g., for even-aged stands, old-growth stands with little recruitment of small trees, or early successional stands with few large trees). Additionally, the large number of transformations of diameter used in modeling tree mortality may reflect the fact that the highly nonlinear relationship between diameter and mortality risk is not well described by parametric forms. Problems in capturing the shape of mortality using parametric forms may be compounded by the fact that logistic regressions are also sensitive to overparameterization (Monserud and Sterba 1999) and collinearity (Yang et al. 2003; Karlsson and Norell 2005; Chao et al. 2008); thus, fitting many transformations of the same covariate (diameter) or adding further covariates might be problematic, particularly because key covariates such as annual growth increment may also covary with diameter (King et al. 2006). Reducing the number of collinear predictors is a possible solution to this last problem. For example, Chao et al. (2008) used model comparison to determine the single diameter transformation that most accurately captured the pattern of mortality over size and included only this variant of diameter-related covariates in their final model. However, another problem then arises: parametric forms where only one transformation of diameter enters the model can prove insufficiently flexible to capture the complex pattern of mortality over size (Wyckoff and Clark 2000). Here, our goal is to introduce a model of mortality built around growth and diameter that addresses many of the problems listed above.

We start from the principle that different processes underlie how the probability of mortality changes with growth and diameter. The increased risk of mortality with decreased growth rate reflects how growth rate integrates resource limitation and other stresses. By contrast, the increased risk of mortality with increased diameter reflects the outcome of structural weakness and senescence that trees display at large size. A key goal of our model is to separate these two processes. Forest simulators often use growth increment as the primary driver of mortality (e.g., JABOWA, Botkin et al. 1972; SORTIE, Pacala et al. 1996). In this framework, higher mortality of large trees can emerge from the fact that growth declines with diameter. Our model allows death of large trees to be considered independently from their declining growth rate.

To capture what is known about the two different processes linking death to growth rate and death to diameter and to avoid making specific assumptions about the shapes of the two relationships (since these are known to be very diverse and highly nonlinear; Wyckoff and Clark 2000), our method links mortality nonparametrically but monotonically to diameter (increasing) and diameter increment (decreasing). This allows for flexibility in the exact relationship between the two covariates and mortality and also forces biological realism to be maintained through the monotonicity. The monotonicity is defined by using a Bayesian framework and establishing priors on the response pattern (Clark 2007, p. 233).

The Bayesian framework also allows inclusion of prior information on the magnitude as well as the direction of the response to each of the covariates. Because large trees are rare in most data sets, including this prior information can be useful in creating realistic models based on limited data. Placing prior weights on estimates allows previous information to be assimilated into inference and facilitates that assimilation of future information (Clark and Gelfand 2006).

The objectives of our study were (*i*) to build a nonparametric Bayesian mortality model and (*ii*) to discuss the importance of our approach in the context of modeling tree mortality. We start by introducing the model, the associated priors, and the Metropolis–Gibbs sampler used to obtain parameter posterior distributions. We then outline the data and detail methods for simulations used to explore where approaches that do not include prior information or constraints on the patterns of change in mortality over size and diameter may struggle. Patterns of mortality obtained using the Bayesian method for four species groups from the southern Appalachians and Piedmont of North Carolina are presented.

Methods

Model

For individual i at time t, the probability of death, q, is defined as

$$[1] \qquad q_{i,t} = \mu_{D_{i,t}} + \mu_{d_{i,t}} - \mu_{D_{i,t}}\mu_{d_{i,t}}$$

where $\mu_{D_{i,t}}$ is the risk of death associated with being in a particular diameter bin D, taken from $\{\boldsymbol{\mu}_D\}$, a vector the length of the number of diameter bins, and $\mu_{d_{i,t}}$ is the risk of death associated with being in a particular diameter increment bin, d (where diameter increment is defined as the difference between diameter in year $t + 1$ and diameter in year t), likewise taken from a vector $\{\boldsymbol{\mu}_d\}$. The product $\mu_{D_{i,t}}\mu_{d_{i,t}}$ accommodates covariance. The two components of the model above ($\mu_{D_{i,t}}$ and $\mu_{d_{i,t}}$) thus separately capture the two different process that drive mortality: (*i*) an increase in mortality with larger diameters to capture physiological limitations associated with large size and (*ii*) a decrease in mortality with increasing diameter increment, where diameter increment captures resource availability. For the results presented here, we set the probability of mortality for the smal-

lest growth increment bin to be one, but this is open to modification (see Discussion).

The first step in fitting the model is defining the vectors of diameter and diameter increment bins, $\{D\}$ and $\{d\}$. In this application, we chose a common set of bins for all species, selected to have equal width on the $\log_{10}$ scale for comparison. However, the specification of bins is flexible. Both the number of bins and their relative magnitude can be tuned to best capture the patterns of interest. Adaptive bin definitions could also be used, e.g., Wyckoff and Clark (2000) defined bins by the minimal number and widths of bins needed to assure monotonicity.

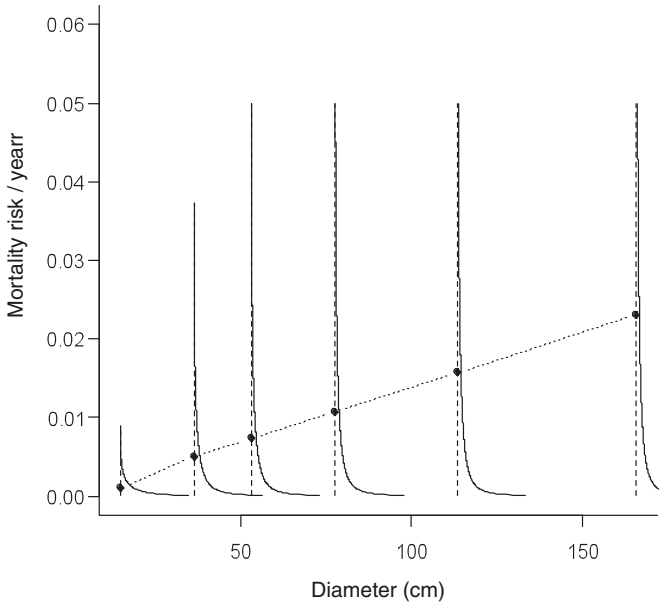
Metropolis algorithm

Having defined the diameter and diameter increment bins, we wish to obtain estimates of the parameter vectors $\{\boldsymbol{\mu}_D\}$ and $\{\boldsymbol{\mu}_d\}$ given the data. The Bernoulli likelihood for the status of individual i at time t, denoted $s_{i,t}$ (where $s_{i,t} = 1$ corresponds to "alive" and $s_{i,t} = 0$ corresponds to "dead"), has mortality probability $q_{i,t}$ calculated from the parameters $\mu_{D_{i,t}}$ and $\mu_{d_{i,t}}$, i.e.:

$$[2] \qquad s_{i,t+1} \sim \text{Bernoulli}(1 - q_{i,t})$$

There are prior constraints that we wish to impose upon $\{\boldsymbol{\mu}_D\}$ and $\{\boldsymbol{\mu}_d\}$. Firstly, we have some prior knowledge about the magnitude of mortality probabilities for particular mortality diameter bins, which is valuable because data may be limited for large diameter trees. This can be captured by defining the prior distribution $\boldsymbol{\mu}_D \sim \text{Beta}(a_D,b_D)$, where a_D and b_D are prior parameters chosen to reflect the expected distributions. For simplicity, we used the same prior for all species groups considered (illustrated in Fig. 1), but different models could be used if more information were available. For prior parameters chosen for the current implementation, the prior median increases with diameter; however, the highest probability density is associated with very low mortality risk, and overall, the prior probability density has a broad distribution, indicating a weak prior (Fig. 1). Similarly, we defined a weak prior for mortality across diameter increment $\boldsymbol{\mu}_d \sim \text{Beta}(a_d,b_d)$. Secondly, we wish to impose monotonicity on the two processes (mortality over diameter and diameter increment) to capture the biological processes outlined in the introduction. This can be achieved when sampling from the posterior density of the parameter by using an indicator function, I(argument),

Fig. 1. Prior density function of mortality risk associated with each diameter bin. The density is oriented towards the y axis (mortality probability), so that the long tails of the beta distribution extend to high mortality risk; the modes of these densities lie along the x axis. The medians for the densities are marked with a small black circle. This figure corresponds to prior parameters of $a_D = 0.5$ and b_D scaled to obtain a chosen mean mortality risk for each bin. Here, for each of the diameter bins, in order of increasing diameter, b_D = 499.5, 99.2, 67.6, 46.0, 31.2, and 21.1.



which takes a value of one when its argument is "true" and zero otherwise (further described below).

We use a Metropolis algorithm to sample from the posterior density, which is the product of the prior density (reflecting constraints or prior beliefs) and the likelihood, reflecting the link between the model and the data. We initiate a Metropolis algorithm with parameter values that monotonically increase $\{\boldsymbol{\mu}_D\}$ or decrease $\{\boldsymbol{\mu}_d\}$ between zero and one. A new sequence of values that are also monotonic and bounded by 0,1 are proposed from normal distributions centered on current parameter values and with a small variance. We denote these candidate parameter values $\{\boldsymbol{\mu}_d^*\}$ and $\{\boldsymbol{\mu}_D^*\}$. An acceptance probability is evaluated for both sets of parameters as

$$[3] \qquad \alpha = \frac{\left(\prod_{i,t} \text{Bernoulli}(s_{i,t}|1 - q_{i,t}^*)\right)\text{Beta}(\mu_D^*|a_D, b_D)\text{Beta}(\mu_d^*|a_d, b_d)}{\left(\prod_{i,t} \text{Bernoulli}(s_{i,t}|1 - q_{i,t})\right)\text{Beta}(\mu_D|a_D, b_D)\text{Beta}(\mu_d|a_d, b_d)} I(\mu_{D-1} < \mu_D^* < \mu_{D+1}) I(\mu_{d-1} > \mu_d^* > \mu_{d+1})$$

where $q_{i,t}^*$ is obtained based on eq. 1 with candidate parameter values $\{\boldsymbol{\mu}_d^*\}$ and $\{\boldsymbol{\mu}_D^*\}$, a_d, b_d, a_D, and b_D are priors set on the distribution of mortality across diameter and diameter increment (see Fig. 1 for an illustration for diameter), and $I(\ )$ is the indicator and is assigned a value one when its argument is true and of zero otherwise. Parameters are updated by the standard Metropolis criterion: if $\alpha > 1$, the proposal is accepted; if not, it is accepted based on α. Incorporating $I(\ )$ into the expression for α ensures that no candidates that violate the prior of monotonicity will be accepted. We repeat this process until the chains of parameter estimates for each parameter converge (Chib and Greenberg 1995), proposing only values that satisfy the indicator to satisfy the monotonicity constraint. All analyses were per-

formed using the software R (R Development Core Team 2007).

Data

We applied the model above to growth and mortality data from nine mapped stands in the southern Appalachians and Piedmont of North Carolina that were established since 1991 for purposes of monitoring and experimental analysis of forest dynamics (Clark et al. 2007; Ibáñez et al. 2007). Five major eastern forest types are represented in the region: mixed hardwoods and eastern hemlock (*Tsuga canadensis* (L.) Carr.) – hardwoods in mesic sites, northern red oak (*Quercus rubra* L.) and chestnut oak (*Quercus prinus* L.) on slopes, white oak (*Quercus alba* L.) and pignut hickory (*Carya glabra* (Mill.) Sweet) on north faces of hilltops, and pitch pine (*Pinus rigida* Mill.) on the southern faces (Whittaker 1956). At the first site, Coweeta Hydrological Laboratory (35°03′N, 83°27′W), where five of the mapped stands are located, all five of these forest types are present. Altitude ranges from 775 to 1480 m, mean soil moisture (or percentage of water content) ranges between 17.8% and 31%, air temperature ranges between 18.4–21.4 °C in July and 0.4–3.5 °C in January (Ibáñez et al. 2007). At the second site, Duke Forest (36.0°N, 79.0°W), where the remaining stands are located, much of the surface corresponds to lands that were abandoned from agriculture early in the last century. Soils are infertile Ultic Alfisols, which are widespread in the Piedmont of North Carolina (Mohan et al. 2006). Mean monthly temperatures range from 19.2 to 31.7 °C in July and from 2.3 to 10.2 °C in December. Approximately 1180 mm of precipitation falls annually. At both sites, previous harvests have limited the number of large trees; few are >100 cm in diameter.

For all trees within each mapped stand with a height >2 m, diameter is repeatedly measured at breast height, which is marked by a nail that holds a tag indicating the identifying tree number (Clark et al. 2007). The complete diameter range available here is 0.12–103 cm. Increment cores were obtained from a subset of trees in 1998 (Wyckoff and Clark 2002) and in 2005. Diameter and diameter increment information are combined in a model to obtain inference on the mean growth increment (the difference between diameter in year $t + 1$ and diameter in year t) and diameter for every individual tree in every year, which are described in detail in Clark et al. (2007). In years where censuses are made, status (alive or dead) is also recorded and verified in subsequent years.

The input to our modeling is the mean of the posterior of growth increment defined for every tree in every year using techniques defined in Clark et al. (2007): the mean posterior of diameter defined for every tree in every year likewise, and the status of each tree (alive or dead) associated with this. Here, we focused on two genera and two species from the full data set: *Acer* spp., *Ilex* spp., *Liriodendron tulipifera* L., and *Liquidambar styraciflua* L., chosen as representative of a range of mortality patterns.

Evaluating the performance of other models

To place our estimates within the context of previous work on tree mortality, we compare the nonparametric Bayesian model we present here with the most common current model, i.e., logistic regression. We were particularly interested in whether or not our model could capture the potentially strong nonlinearity between growth and mortality and diameter and mortality and if we could parameterize the contribution of growth as separate from size, and how its performance compared with standard parametric models. Accordingly, we fit two separate logistic regressions: one including diameter (D) and diameter increment (d) and one including the inverse of diameter (D^{-1}) and diameter increment, which is in line with previous work on tree mortality. To fit these models, we used the general linear methods (GLM) function in R (R Development Core Team 2007). We provide parameter values and the log-likelihood from our nonparametric mortality model and the logistic regressions.

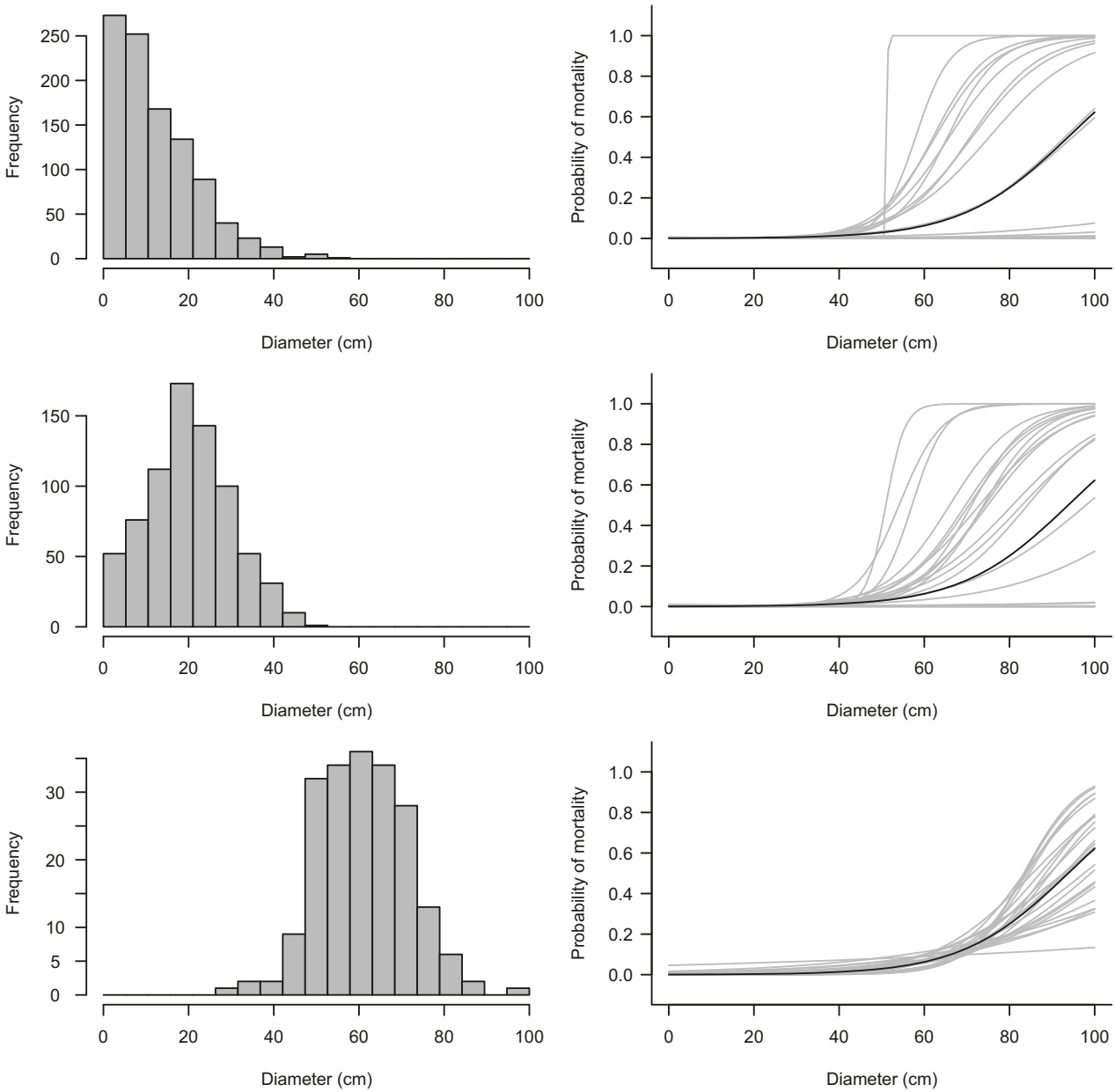
To demonstrate the sensitivity of logistic regressions to tree size distributions and to highlight the importance of the prior on the magnitude of mortality at large-diameter bins, we also generated three different distributions for diameter and simulated mortality according to the same true pattern of mortality over size for all three, which is shown as the black line in Fig. 2, second column of graphs. This true pattern was used to simulate 25 sets of mortality data for the size distribution in the first column and logistic regressions were used to test recovery of the true patterns over size.

Slow-growing trees are likely to be consistently slow growing. If slow growth is associated with low survival probability, this will strongly bias the distribution of growth increments upwards with the passage of time because slow-growing trees may survive a few years but are unlikely to survive longer, so that after a few years only fast-growing trees will remain. This may affect the fit of parametric models like the logistic regression. To explore this, we set a highly nonlinear relationship between growth and mortality (following the results of Kobe and Coates (1997) and Wyckoff and Clark (2002); see Fig. 3*b*, black solid line) and initiated a population of individuals with an even distribution of diameter increment (Fig. 3*a*, black bars). We then generated deviates from a binomial distribution with probability set by the chosen probability of survival associated with each diameter increment taken to the power of 10 to identify individuals that would survive for 10 years (Fig. 3*a*, grey shaded bars). We then generated mortality data for (*i*) the initial even distribution of diameter increments and (*ii*) the distribution of diameter increments obtained after 10 years using a binomial distribution and the "true" probability of mortality distribution. We fitted logistic regressions to both using the GLM function as above, with D and D^{-1} as covariates to maximize the chances of the model of capturing the fit, and explored the degree to which distributional bias is likely to result from the types of dynamics relevant for growth increment might affect model fitting.

Results

Fitted models and data are shown in Fig. 4. For all four examples, the inference on the pattern of declining mortality with increasing diameter increment is precise (the posteriors indicated by broken lines are narrow; Figs. 3 and 4). The pattern of increasing mortality with increasing diameter is less confident (the posteriors are wider), which reflects the

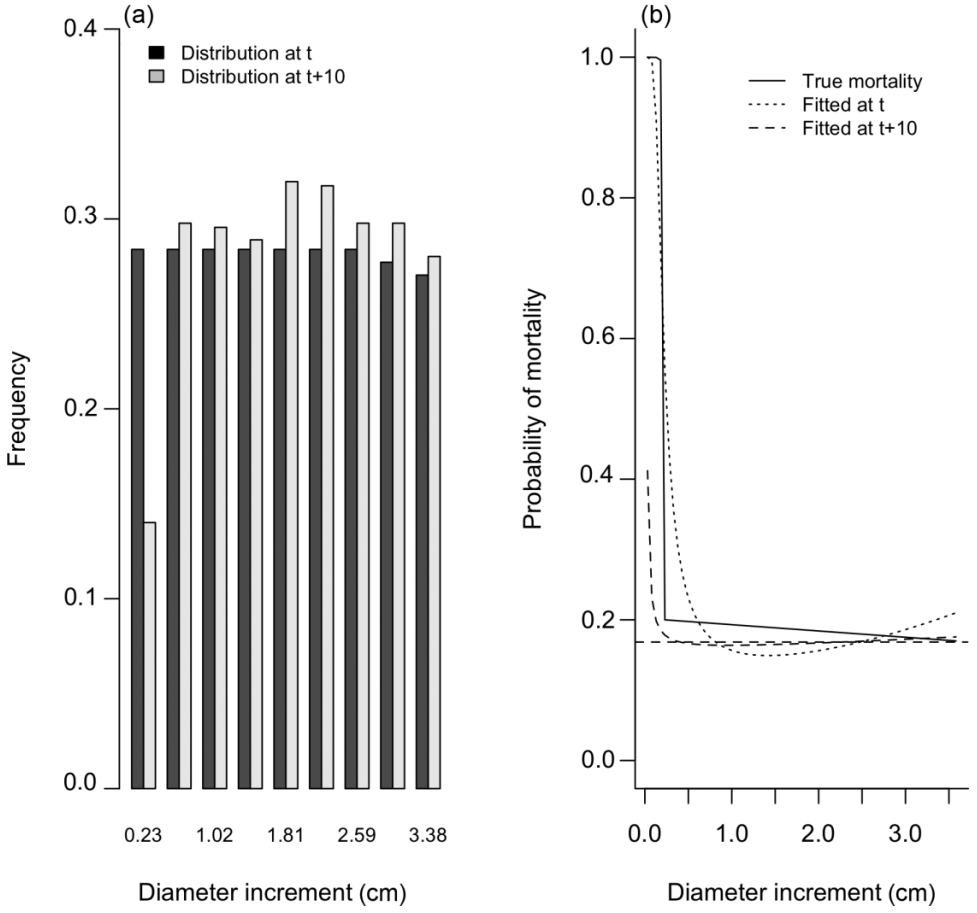
Fig. 2. Simulated diameter distributions for three successional stages. The first column shows the size distributions of individual trees corresponding to early succession (first row, very few large trees), midsuccession (second row, a shift to midsized trees), and late succession with very few small recruits (third row). The number of individual stems is 1000, 750, and 200 over the three forest types, and their estimated basal area (assuming a 1 ha stand) is 20, 30, and 60 m^2/ha). The second column shows a chosen "true" underlying pattern of increasing mortality probability with diameter (black line, the same for all three cases), chosen so that mortality gradually rises after a diameter of 40 cm (similar to *Liquidambar styraciflua*; see Fig. 1). Results of logistic regressions fitted to simulated data sets are shown in grey.



paucity of observations at this diameter range (Fig. 4), with most information coming from the prior. The data (shown in Fig. 4) are not overwhelmed by our prior specification on the diameter sequence (Fig. 1), which is informative at values near zero for small diameter classes but less informative for large diameter classes. Overall, the patterns reflect how informative priors, including monotonicity, contribute to inference. For example, although many small individuals die (Fig. 4, histograms, right-hand column), the mortality risk is low for low diameters (Fig. 4, lines, right-hand column). However, smaller individuals tend to have small diameter growth increments, so that these observed deaths are captured by the high mortality risk of slow-growing individuals (Fig. 4, left-hand column, lines).

In comparing our new model with logistic regressions, the negative log-likelihoods are lower for the latter (Table 1). This is expected because we have used informative priors to obtain a fit that balances data and prior, as opposed to finding the fit that maximizes the likelihood alone (the posterior of the Bayesian model multiplies the likelihood by the prior and a log-likelihood test cannot capture this interaction). Our nonparametric fit assumes that zero growth, at least

Fig. 3. (*a*) The even frequencies of diameter increment, *d*, taken at the start of the simulation before differential mortality sampling occurs (black bars); and the biased frequencies after 10 years, with fewer small *d* (grey bars). Values on the *x* axis correspond to midpoints of the bins used for the histograms. (*b*) The "true" relationship between diameter increment and mortality (black solid line) used to obtain the distribution of diameters after 10 years (grey bars in Fig. 3*a*) and fitted curves for mortality data simulated from this "true" relationship from (*i*) the even distribution at the start of the simulation (black dotted line, corresponding to black bars in Fig. 3*a*, both *d* and d^{-1} are significant) and (*ii*) the distribution of diameter increment that would have resulted if individuals retained the same diameter increment for 10 years, and thus, the slowest growing individuals had low survival and had been selected out (curved black broken line corresponding to grey bars in Fig. 3*a*; this fit had a nonsignificant *d* and d^{-1} effect; the horizontal black broken line shows a fit with only the intercept).

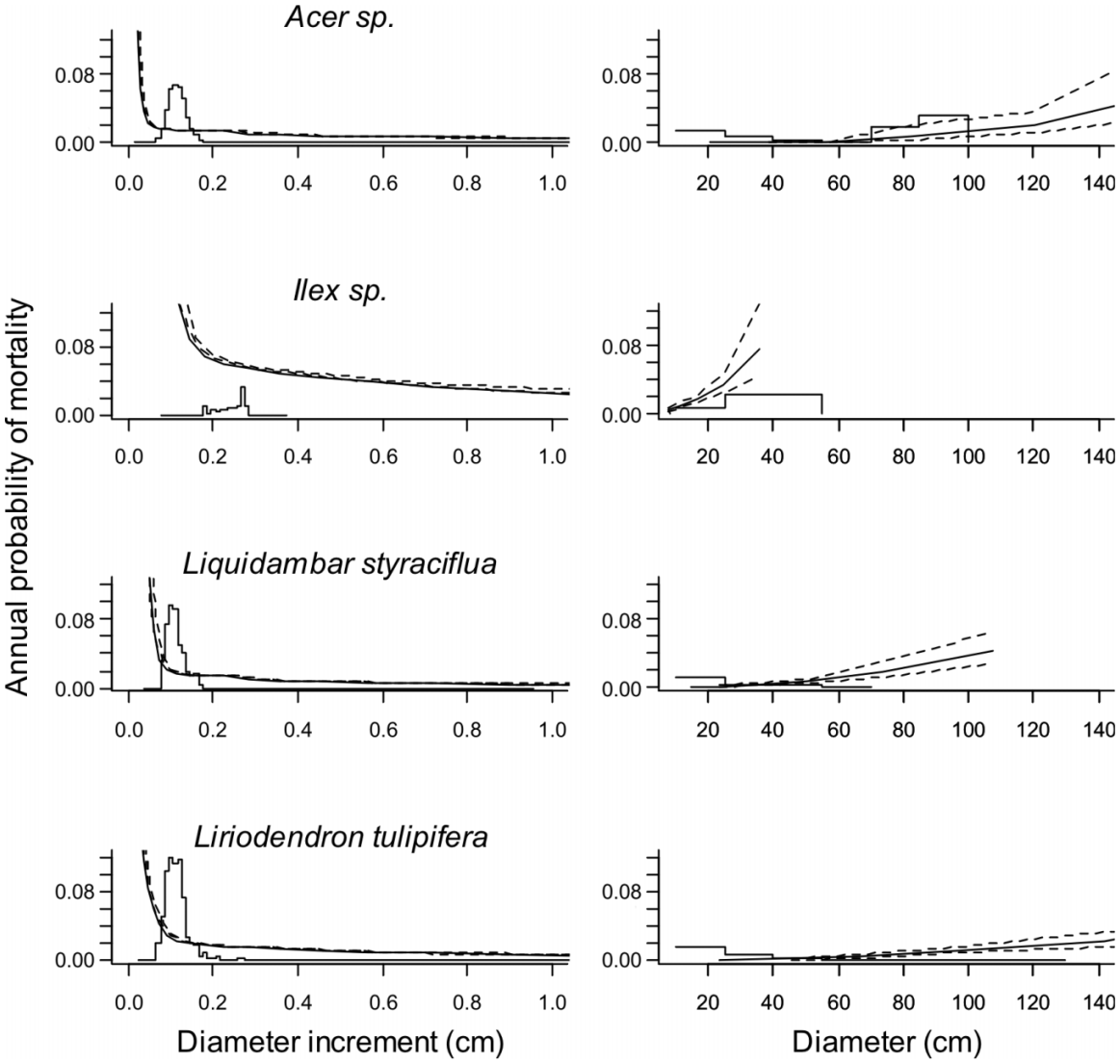


eventually, is consonant with death (e.g., the value of μ_d is 0 for the smallest diameter increment bin), whereas a logistic regression will accept unreasonably small intercepts for mortality risk at low growth rate. This intercept cannot be well informed by data because high mortality risk at low growth rate assures that few observations occur here (see below). Furthermore, coefficients of parametric logistic regressions are largely nonsignificant and indicate variable patterns over diameter and diameter increment (Table 1, Fig. 5). In fact, for diameter increment, the pattern contradicts that which we know to be operating (faster growth appears to lead to higher mortality instead of the reverse).

Where the data are restricted to certain size ranges, as occurs in actual data sets, the inference on patterns of mortality across diameter can vary substantially for logistic regressions and shows considerable instability (Fig. 2). With few large trees, even with a sample size of 1000, the prediction made by simulations using logistic regression is either of no increase of mortality with diameter (grey lines in Fig. 2, first row) or an extremely steep increase. With more large trees, even as sample sizes decline to 750 and 200 (as would happen in a maturing forest) (Fig. 2, second and third row), the mean mortality trajectory is closer to the true trajectory, but the range of patterns observed is considerable (Fig. 2, second row). This indicates that the number of trees used to parameterize a logistic model of mortality due to diameter is not as important as the number of large trees. This model has no place to include information about large tree mortality unless that information is in the data at hand. Our nonparametric Bayesian approach can use knowledge gleaned from other studies of large-tree mortality for application to mortality studied in any forest.

Where logistic regressions are fitted before and after the sampling effect of 10 years of survival is applied to diameter increment, the bias in distribution of diameter increment resulting from sampling through survival can likewise lead to patterns that do not capture a true relationship between diameter increment and mortality (Fig. 3). Before the sampling occurs, the best-fitted logistic regression has a linear predictor of the form $2.55 - 0.58D^{-1} - 0.25D$, with all coefficients significantly different from zero ($P < 0.01$), and the curve corresponds quite closely to the true relationship

Fig. 4. The two components of the nonparametric mortality model for the two species and genera considered. The fitted model (solid line) with 95% credible intervals (broken lines) are shown as are the observations (histograms of diameter and diameter increment of individuals that die).



(Fig. 3*b*, solid line compared with the dotted lines). After sampling has occurred and there are few slow-growing individuals (Fig. 3*a*, grey bars, overall $\sim$1000 of the 1500 initial individuals remain, and loss is concentrated in the smallest size class), the logistic regression linear predictor is $1.23 + 0.07D^{-1} + 0.09D$ (Fig. 3*b*, curved broken line). Only the intercept is significantly different from zero, so that no effect of diameter increment could be concluded (Fig. 3*b*, horizontal broken line). Whether the effect of diameter is retained or not, fitted logistic regressions can miss high mortality of slow-growing individuals.

Discussion

By separating the effects of growth increment from the effects of diameter on the risk of mortality, we obtained estimates of mortality risks associated with low vigor and large size that reflect knowledge (*i*) that these relationships are monotonic, (*ii*) that high mortality of small individuals reflects vigor and not size, and (*iii*) that the relationships can be more nonlinear than can be approximated with simple parametric forms. *Liriodendron tulipifera* and *Liquidambar styraciflua*, which are typically considered shade intolerant, are more sensitive to low growth rates than the shade-tolerant *Acer* spp. This result aligns with previous work indicating that shade-tolerant species are more likely to survive the resource limitation conditions captured by suppressed growth (Kobe and Coates 1997). Western redcedar (*Thuja plicata* Donn.), which was the most shade-tolerant species in Kobe and Coates (1997), attained indefinite survival once growth increment was greater than 0.03 cm/year. The most shade-tolerant species group in our results, *Acer*, reaches a comparable level of survival at a growth increment of 0.05 cm/year. For a cohort of individuals growing at this rate, it would take about 100 years for the proportion of survivors to drop below 10% (Fig. 5). By contrast, in the same growth conditions (0.05 cm/year), the least tolerant species in our results, *Ilex decidua* Walt., will drop under

Table 1. Linear predictor associated with the logistic regressions, parameter standard errors, and log likelihood for all three different models.

		Parameter SEs			
	Linear predictors	Intercept	D or D^{-1}	d	Log likelihood
***Acer* sp. (n_{sp} = 59 863, n_{dead} = 696)**					
Logistic regression (D, d)	$-4.48 + 0.002D + 0.15d$	0.093***	0.004	0.468	–3797
Logistic regression (D^{-1}, d)	$-4.46 - 0.013D^{-1} + 0.18d$	0.112***	0.070	0.486	–3798
Nonparametric Bayes	—				–4534
***Ilex* sp. (n_{sp} = 8686, n_{dead} = 69)**					
Logistic regression (D, d)	$-4.59 + 0.03D - 1.66d$	1.075***	0.026	4.822	–407
Logistic regression (D^{-1}, d)	$-4.09 - 1.35D^{-1} - 1.22d$	1.066***	0.630**	4.768	–405
Nonparametric Bayes	—				–858
***Liquidambar styraciflua* (n_{sp} = 14 717, n_{dead} = 155)**					
Logistic regression (D, d)	$-4.50 - 0.01D + 0.49d$	0.190***	0.009	0.781	–864
Logistic regression (D^{-1}, d)	$-4.61 + 0.04D^{-1} + 0.26d$	0.211***	0.170*	0.787	–865
Nonparametric Bayes	—				–1057
***Liriodendron tulipifera* (n_{sp} = 8986, n_{dead} = 109)**					
Logistic regression (D, d)	$-4.55 + 0.001D + 0.40d$	0.178	0.005	0.489	–594
Logistic regression (D^{-1}, d)	$-4.53 - 0.008D^{-1} + 0.44d$	0.216	0.166	0.498	–594
Nonparametric Bayes	—				–721

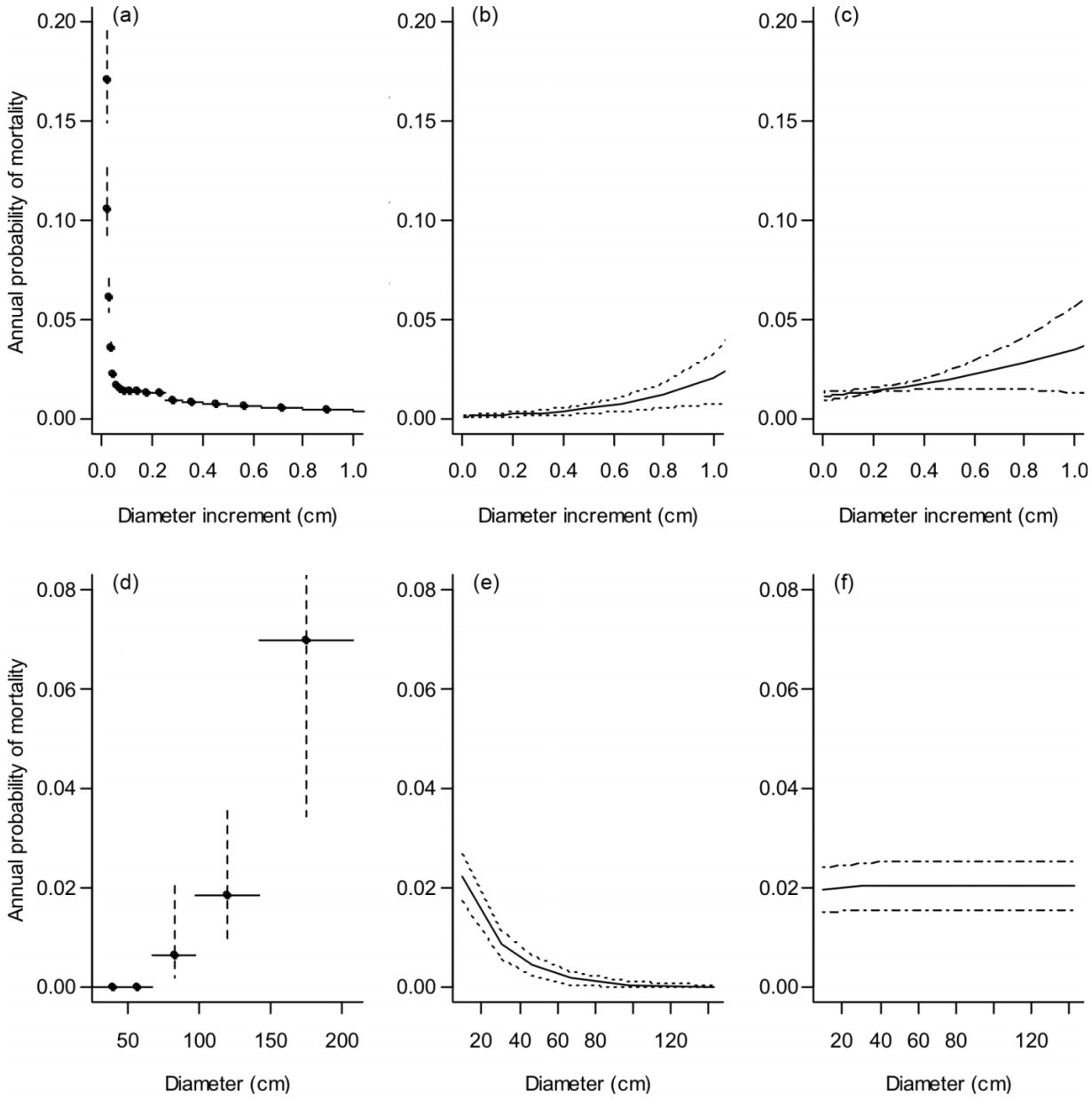
Note: To obtain the probability of mortality, the linear predictor x was transformed by $\exp(x)/(1 + \exp(x))$. The SEs of the linear predictors that are significantly different from zero are given with asterisks as follows: *, $P < 0.1$; **, $P < 0.05$; ***, $P < 0.0001$. n_{sp}, total number of indivduals of the species or species group; n_{dead}, number of dead individuals in the species or species group.

10% survivorship in only 1 year. *Ilex decidua* also has the fastest increasing mortality with increasing diameter size. *Ilex decidua* is generally a subcanopy species, with a much smaller maximum diameter than the other species (Fig. 4), so that structural features capable of preventing increasing mortality at large sizes may be under less strong selection in this species, leading to the rapid increase of mortality with diameter. *Liriodendron tulipifera*, which can attain considerable heights, has the slowest increase in mortality with diameter, compatible with long-term structural strength. Monserud and Sterba (1999) found that, for spruce, mortality first declined to a probability below 0.02 with increasing diameter; however, after a diameter of approximately 70 cm, mortality more than doubled, reaching a probability of 0.05 per year. For all species groups we consider here, the increase in mortality starts before 70 cm, and accelerates more rapidly than recorded by Monserud and Sterba (1999). The structurally driven increase of mortality of large trees may be mitigated by an increase in growth rate that large trees can experience as their light exposure increases. Our model will take this into account and separate the effects of diameter from the effects of growth. This may explain why our increase in mortality starts at a much smaller diameter.

How does this model fit into the context of previous modeling of mortality, and what are its benefits? Likelihood comparisons appear to undermine the value of the nonparametric Bayesian model we present relative to logistic regressions (Table 1). However, although the log-likelihood may be lower, the nonparametric Bayesian approach is explicitly weighted by the prior as well as the data, so comparison to the data alone will inevitably be weakened. We are formally incorporating information not directly in the data but available from other sources, i.e., the known increase of mortality at large sizes and the known decrease of mortality with increasing growth rate. In other words, the Bayesian approach "sacrifices" model fit to any one data set to balance the contribution from data and prior. The advantage of an informative prior always has this effect of weakening the fit to the data. This sacrifice is desirable because it allows the incorporation of a minimally constraining prior (monotonicity), which is known to be true, without specifying a particular functional form. This is of particular value when data are sparse, as they will be across tree diameter and diameter increment samples. Despite large overall sample sizes for the four species considered here, the effect of diameter increment in logistic regressions here is often negative and, overall, is not significant; the effect of diameter is also often not significant (Table 1). For *Acer*, the logistic regression with D predicts the reverse of the expected pattern of mortality across diameter (Fig. 1). The logistic regression with D^{-1} captures increasing mortality risk with increasing diameter but, as for the model with D, fails to capture very high mortality of very slow-growing trees. If only significant parameters were used, there would be no effect of diameter or diameter increment (Table 1). Other species groups show similar results.

Any parametric methods (e.g., logistic regression) will be sensitive to the distribution of data. This will adversely influence inference at diameters and growth increments where mortality is rarely observed (Lavine 1991). For example, there are generally very few data at the smallest end of the growth increment scale (e.g., see histograms in Fig. 1), where mortality is often found to be highly nonlinear (Kobe and Coates 1997; Wyckoff and Clark 2002). The lack of observations of individuals with very slow growth is presumably a result of their rapid mortality and disappearance, leaving a nonrandom sample of better surviving individuals. However, the fitted parametric model will always be dominated by the abundant growth rates at higher levels of growth and, consequently, will predict low risks of mortality

Fig. 5. For *Acer*, the predicted probability of mortality per year (*y* axis) over diameter increment (first row) and diameter (second row), showing a comparison of the three fitted models with credible and confidence intervals: (*a*) and (*d*) the Bayesian nonparametric model; (*b*) and (*e*) the logistic regression with D and d; and (*c*) and (*f*) the logistic regression with D^{-1} and d. For the logistic regressions, mortality cannot be calculated in the absence of one of the two covariates, so we used $D = 40$ and $d = 0.5$ (altering this does not alter patterns).



for the smallest growth rates. The simulation illustrated in Fig. 3 shows that this failure will be particular evident where the relationship is highly nonlinear at lower growth rates (e.g., Kobe and Coates 1997; Wyckoff and Clark 2002).

Kobe and Coates (1997) found that the intercept of their mortality function rarely differed significantly from 1. In line with this, here we set the probability of mortality for individuals with a zero growth rate to 1, but it is important to recognize that this does not always necessarily hold. Some species are capable of suspending radial growth, so that missing rings may result (Lorimer et al. 1999), and zero growth rates are not necessarily associated with mortality. However, this assumption of our model is easily modified, either by changing the intercept to be less than 1 (and this could be directly fitted) or by using a time step of more than 1 year. Further, for species that can sustain very low growth for many years, smaller bin sizes could be used at the slow growth end of the scale to pick up the declines in growth rates beyond the very slow baseline, which are indicative of resource stress and, therefore, increased probability of mortality.

Our model offers an important, flexible extension of mortality modeling. A range of models has been developed for use where only population-level data are available (Lieberman et al. 1985; Korning and Balslev 1994; van Mantgem and Stephenson 2007); for our model, at a minimum, individual diameter, growth increment, and status in-

formation are necessary. Although our model is geared towards studies of forest stand dynamics, large-scale (continent-wide) simulations of forest dynamics will likely face similar data-restriction problems. Rare species, which occur in any forest data set, will have few data with which to estimate mortality rates. In common with binomial models used by King et al. (2006), our model uses discrete size classes but deviates from these models chiefly through added monotonicity. Our model focuses only on living trees (unlike Kobe et al. 1995; Kobe and Coates 1997; Wyckoff and Clark 2000); however, Wyckoff and Clark (2002) use a similar nonparametric form (although applied to living and dead trees), but bin widths are modeled as adaptive so that the optimal structure of bins is developed simultaneously with estimates of mortality probabilities associated with each bin. This could also be incorporated for the model we present here; however, to facilitate comparisons among species groups, we chose to retain the same structure for all four in this instance.

Many previous models of mortality have used several other covariates beyond diameter and diameter increment (e.g., crown ratio, crown defoliation, conspecific density, stand density, stand age, individual age, basal area per hectare, basal area in larger trees, site productivity index, and wind intensity). If further covariates were considered important, these could be accommodated simply by adding terms to eq. 1. However, by using only growth and size as covariates, our model lends itself more directly to implementation in a simulator, because any other covariates that were fitted statistically would also need simulating. Using growth as a proxy for other covariates is also a valid approach, because growth is expected to integrate factors ranging from stress, competition, and environmental variation with individual effects, which is a key component of confidence in inferred mortality patterns due to sampling issues that might otherwise dominate mortality estimates (Vaupel et al. 1979). Estimating individual variation in mortality probability is challenging because individuals only die once, but using individual variation in growth as a key covariate allows variance components to be estimated in a straightforward manner. Likewise, annual effects (e.g., van Mantgem and Stephenson 2007) can be captured by the growth model on which our mortality model is built (Clark et al. 2007). If the relationship between growth and mortality changes across sites with different soil or altitudinal characteristics (e.g., Kobe 1996), different models could be fitted in different sites. A potentially important complication is that changes in growth rather than growth itself may be the key covariate of probability of mortality. For example, Bigler and Bugmann (2004) found that a sudden decline in growth is a reliable indicator of an increase in mortality probability in spruce species. An eventual extension of the model we introduce here would be to include another component capturing the short-term change in growth.

To conclude, we have proposed a flexible nonparametric form for modeling tree mortality that lends itself to transparent inclusion of priors and has proved capable of capturing a range of mortality trajectories in a manner consistent with physiological expectations. The benefits of including such biological realism may be particularly valuable where mortality models are integrated into forest simulators. Mortality of large trees will not be neglected even where data are sparse, and inference may more realistically reflect patterns of mortality over ontogeny than inference obtained through unconstrained parametric logistic regressions.

Acknowledgements

This manuscript was substantially improved through comments by two anonymous reviewers.

References

Bigler, C., and Bugmann, H. 2004. Predicting the time of tree death using dendrochronological data. Ecol. Appl. **14**(3): 902–914. doi:10.1890/03-5011.

Botkin, D.B., Janak, J.F., and Wallis, J.R. 1972. Some ecological consequences of a computer model of forest growth. J. Ecol. **60**(3): 849–872.

Burleigh, J.S., Alfaro, R.I., Borden, J.H., and Taylor, S. 2002. Historical and spatial characteristics of spruce budworm *Choristoneura fumiferana* (Clem.) (Lepidoptera: Tortricidae) outbreaks in northeastern British Columbia. For. Ecol. Manage. **168**(1–3): 301–309. doi:10.1016/S0378-1127(01)00748-4.

Caspersen, J.P., Pacala, S.W., Jenkins, J.C., Hurtt, G.C., Moorcroft, P.R., and Birdsey, R.A. 2000. Contributions of land-use history to carbon accumulation in U.S. forests. Science (Washington, D.C.), **290**(5494): 1148–1151. doi:10.1126/science.290.5494.1148. PMID:11073451.

Chao, K.J., Phillips, O.L., Gloor, E., Monteagudo, A., Torres-Lezama, A., and Martínez, R.V. 2008. Growth and wood density predict tree mortality in Amazon forests. J. Ecol. **96**(2): 281–292. doi:10.1111/j.1365-2745.2007.01343.x.

Cherubini, P., Fontana, G., Rigling, D., Dobbertin, M., Brang, P., and Innes, J.L. 2002. Tree-life history prior to death: two fungal root pathogens affect tree-ring growth differently. J. Ecol. **90**(5): 839–850. doi:10.1046/j.1365-2745.2002.00715.x.

Chib, S., and Greenberg, E. 1995. Understanding the Metropolis–Hastings algorithm. Am. Stat. **49**(4): 327–335. doi:10.2307/2684568.

Clark, D.A. 2002. Are tropical forests an important carbon sink? Reanalysis of the long-term plot data. Ecol. Appl. **12**(1): 3–7. doi:10.1890/1051-0761(2002)012[0003:ATFAIC]2.0.CO;2.

Clark, D.A. 2004. Sources or sinks? The responses of tropical forests to current and future climate and atmospheric composition. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **359**(1443): 477–491. doi:10.1098/rstb.2003.1426.

Clark, J.S. 2007. Models for ecological data. Princeton University Press, Princeton, N.J.

Clark, D.A., and Clark, D.B. 1999. Assessing the growth of tropical rain forest trees: issues for forest modeling and management. Ecol. Appl. **9**(3): 981–997. doi:10.1890/1051-0761(1999)009[0981:ATGOTR]2.0.CO;2.

Clark, J.S., and Gelfand, A.E. 2006. A future for models and data in environmental science. Trends Ecol. Evol. **21**(7): 375–380. doi:10.1016/j.tree.2006.03.016. PMID:16815437.

Clark, J.S., Wolosin, M., Dietze, M., Ibáñez, I., LaDeau, S., Welsh, M., and Kloeppel, B. 2007. Tree growth inference and prediction from diameter censuses and ring widths. Ecol. Appl. **17**(7): 1942–1953. doi:10.1890/06-1039.1. PMID:17974333.

Condit, R., Hubbell, S.P., and Foster, R.B. 1993. Identifying fast-growing native trees from the neotropics using data from a large, permanent census plot. For. Ecol. Manage. **62**(1–4): 123–143. doi:10.1016/0378-1127(93)90046-P.

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. Ecol. Monogr. **65**(4): 419–439. doi:10.2307/2963497.

Condit, R., Ashton, P., Bunyavejchewin, S., Dattaraja, H.S., Davies, S., Esufali, S., Ewango, C., Foster, R., Gunatilleke, I.A.U.N., Gunatilleke, C.V.S., Hall, P., Harms, K.E., Hart, T., Hernandez, C., Hubbell, S., Itoh, A., Kiratiprayoon, S., Lafrankie, J., de Lao, S.L., Makana, J.-R., Noor, M.N.S., Kassim, A.R., Russo, S., Sukumar, R., Samper, C., Suresh, H.S., Tan, S., Thomas, S., Valencia, R., Vallejo, M., Villa, G., and Zillio, T. 2006. The importance of demographic niches to tree diversity. Science (Washington, D.C.), **313**(5783): 98–101. doi:10.1126/science.1124712. PMID:16763113.

Coomes, D.A., and Allen, R.B. 2007. Mortality and tree-size distributions in natural mixed-age forests. J. Ecol. **95**(1): 27–40. doi:10.1111/j.1365-2745.2006.01179.x.

Eid, T., and Tuhus, E. 2001. Models for individual tree mortality in Norway. For. Ecol. Manage. **154**(1–2): 69–84. doi:10.1016/S0378-1127(00)00634-4.

Gilbert, B., Wright, S.J., Muller-Landau, H.C., Kitajima, K., and Hernandéz, A. 2006. Life history trade-offs in tropical trees and lianas. Ecology, **87**(5): 1281–1288. doi:10.1890/0012-9658(2006)87[1281:LHTITT]2.0.CO;2. PMID:16761606.

Ibáñez, I., Clark, J.S., LaDeau, S., and Lambers, J.H.R. 2007. Exploiting temporal variability to understand tree recruitment response to climate change. Ecol. Monogr. **77**(2): 163–177. doi:10.1890/06-1097.

Karlsson, K., and Norell, L. 2005. Modelling survival probability of individual trees in Norway spruce stands under different thinning regimes. Can. J. For. Res. **35**(1): 113–121. doi:10.1139/x04-161.

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. For. Ecol. Manage. **223**(1–3): 152–158. doi:10.1016/j.foreco.2005.10.066.

Kobe, R.K. 1996. Intraspecific variation in sapling mortality and growth predicts geographic variation in forest composition. Ecol. Monogr. **66**(2): 181–201. doi:10.2307/2963474.

Kobe, R.K., and Coates, K.D. 1997. Models of sapling mortality as a function of growth to characterize interspecific variation in shade tolerance of eight tree species of northwestern British Columbia. Can. J. For. Res. **27**(2): 227–236. doi:10.1139/cjfr-27-2-227.

Kobe, R.K., Pacala, S.W., Silander, J.A., Jr., and Canham, C.D. 1995. Juvenile tree survivorship as a component of shade tolerance. Ecol. Appl. **5**(2): 517–532. doi:10.2307/1942040.

Koch, G.W., Sillett, S.C., Jennings, G.M., and Davis, S.D. 2004. The limits to tree height. Nature (London), **428**(6985): 851–854. doi:10.1038/nature02417. PMID:15103376.

Korning, J., and Balslev, H. 1994. Growth rates and mortality patterns of tropical lowland tree species and the relation to forest structure in Amazonian Ecuador. J. Trop. Ecol. **10**(2): 151–166. doi:10.1017/S026646740000780X.

Lavine, M. 1991. Problems in extrapolation illustrated with space shuttle O-ring data. J. Am. Stat. Assoc. **86**(416): 919–921. doi:10.2307/2290505.

Lieberman, D., Lieberman, M., Peralta, R., and Hartshorn, G.S. 1985. Mortality patterns and stand turnover rates in a wet tropical forest in Costa Rica. J. Ecol. **73**(3): 915–924. doi:10.2307/2260157.

Lorimer, C.G., Dahir, S., and Singer, M.T. 1999. Frequency of partial and missing rings in *Acer saccharum* in relation to canopy position and growth rate. Plant Ecol. **143**(2): 189–202. doi:10.1023/A:1009847819158.

Mohan, J.E., Clark, J.S., and Schlesinger, W.H. 2007. Long-term CO_2 enrichment of a forest ecosystem: implications for forest regeneration and succession. Ecol. Appl. **17**(4): 1198–1212. doi:10.1890/05-1690.

Monserud, R.A., and Sterba, H. 1999. Modeling individual tree mortality for Austrian forest species. For. Ecol. Manage. **113**(2–3): 109–123. doi:10.1016/S0378-1127(98)00419-8.

Nord-Larsen, T. 2006. Modeling individual-tree growth from data with highly irregular measurement intervals. For. Sci. **52**: 198–208.

Pacala, S.W., Canham, C.D., Saponara, J., Silander, J.A., Jr., Kobe, R.K., and Ribbens, E. 1996. Forest models defined by field measurements: II. Estimation, error analysis and dynamics. Ecol. Monogr. **66**(1): 1–43. doi:10.2307/2963479.

Platt, W.J., Evans, G.W., and Rathbun, S.L. 1988. The population dynamics of a long-lived conifer (*Pinus palustris*). Am. Nat. **131**(4): 491–525. doi:10.1086/284803.

R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

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. J. Ecol. **95**(6): 1261–1273. doi:10.1111/j.1365-2745.2007.01301.x.

Ryan, M.G., and Yoder, B.J. 1997. Hydraulic limits to tree height and tree growth. Bioscience, **47**(4): 235–242. doi:10.2307/1313077.

Suarez, M.L., Ghermandi, L., and Kitzberger, T. 2004. Factors predisposing episodic drought-induced tree mortality in *Nothofagus* site, climatic sensitivity and growth trends. J. Ecol. **92**(6): 954–966. doi:10.1111/j.1365-2745.2004.00941.x.

van Mantgem, P.J., and Stephenson, N.L. 2007. Apparent climatically induced increase of tree mortality rates in a temperate forest. Ecol. Lett. **10**(10): 909–916. doi:10.1111/j.1461-0248.2007.01080.x. PMID:17845291.

Vaupel, J.W., Manton, K.G., and Stallard, E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, **16**(3): 439–454. doi:10.2307/2061224. PMID:510638.

Whittaker, R.H. 1956. Vegetation of the Great Smoky Mountains. Ecol. Monogr. **26**(1): 1–80. doi:10.2307/1943577.

Wunder, J., Reineking, B., Matter, J.-F., Bigler, C., and Bugmann, H. 2007. Predicting tree death for *Fagus sylvatica* and *Abies alba* using permanent plot data. J. Veg. Sci. **18**: 525–534. doi:10.1139/cjfr-30-1-156.

Wyckoff, P.H., and Clark, J.S. 2000. Predicting tree mortality from diameter growth: a comparison of maximum likelihood and Bayesian approaches. Can. J. For. Res. **30**(1): 156–167. doi:10.1139/cjfr-30-1-156.

Wyckoff, P.H., and Clark, J.S. 2002. The relationship between growth and mortality for seven co-occurring tree species in the southern Appalachian Mountains. J. Ecol. **90**(4): 604–615. doi:10.1046/j.1365-2745.2002.00691.x.

Yang, Y., Titus, S.J., and Huang, S. 2003. Modeling individual tree mortality for white spruce in Alberta. Ecol. Modell. **163**(3): 209–222. doi:10.1016/S0304-3800(03)00008-5.

Yoda, K., Shinozaki, K., Ogawa, H., Hozumi, K., and Kira, T. 1965. Estimation of the total amount of respiration in woody organs of trees and forest communities. J. Biol. Osaka City Univ. **16**: 15–26.

Zhao, D., Borders, B., and Wilson, M. 2004. Individual-tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississipi alluvial valley. For. Ecol. Manage. **199**(2–3): 307–322. doi:10.1016/j.foreco.2004.05.043.