Inference for Size Demography From Point Pattern Data Using Integral Projection Models

Souparno GHOSH, Alan E. GELFAND, and James S. CLARK

Population dynamics with regard to evolution of traits has typically been studied using matrix projection models (MPMs). Recently, to work with continuous traits, integral projection models (IPMs) have been proposed. Imitating the path with MPMs, IPMs are handled first with a fitting stage, then with a projection stage. Fitting these models has so far been done only with individual-level *transition* data. These data are used to estimate the demographic functions (survival, growth, fecundity) that comprise the *kernel* of the IPM specification. Then, the estimated kernel is iterated from an initial trait distribution to project steady state population behavior under this kernel. When trait distributions are observed over time, such an approach does not align projected distributions with these observed temporal benchmarks.

The contribution here, focusing on size distributions, is to address this issue. Our concern is that the above approach introduces an inherent mismatch in scales. The redistribution kernel in the IPM proposes a mechanistic description of population level redistribution. A kernel of the same functional form, fitted to data at the individual level, would provide a mechanistic model for individual-level processes. Resulting parameter estimates and the associated estimated kernel are at the wrong scale and do not allow population-level interpretation.

Our approach views the observed size distribution at a given time as a point pattern over a bounded interval. We build a three-stage hierarchical model to infer about the dynamic intensities used to explain the observed point patterns. This model is driven by a *latent* deterministic IPM and we introduce uncertainty by having the operating IPM vary around this deterministic specification. Further uncertainty arises in the realization of the point pattern given the operating IPM. Fitted within a Bayesian framework, such modeling enables full inference about all features of the model. Such dynamic modeling, *optimized* by fitting to data observed over time, is better suited to projection.

Exact Bayesian model fitting is very computationally challenging; we offer approximate strategies to facilitate computation. We illustrate with simulated data examples as well as well as a set of annual tree growth data from Duke Forest in North Carolina. A further example shows the benefit of our approach, in terms of projection, compared with the foregoing individual level fitting.

Souparno Ghosh is a post-doctoral researcher (E-mail: *sg147@stat.duke.edu*) and Alan E. Gelfand (✉) is a professor (E-mail: *alan@stat.duke.edu*), Department of Statistical Science, Duke University, Durham, NC, USA. James S. Clark is a professor (E-mail: *jimclark@duke.edu*), Nicholas School of Environment, Duke University, Durham, NC, USA.

This invited paper is discussed in the comments available at doi:*10.1007/s13253-012-0119-5*, doi:*10.1007/s13253-012-0120-z*, doi:*10.1007/s13253-012-0121-y*.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 17, Number 4, Pages 641–677
DOI: 10.1007/s13253-012-0123-9

Key Words: Density dependence; Fourier transform; Hierarchical model; Integro-difference equation; Laplace approximation; Nonhomogeneous Poisson process; Spectral domain.

1. INTRODUCTION

The study of population dynamics has a long history in ecology and biology with dynamics typically summarized by age, size or stage distributions (see, e.g., Keyfitz and Caswell 2005 and references therein). Analysis of changing structure requires a translation of individual level data to the population level. Ecologists and conservation biologists primarily use matrix projection models (MPM) to make this translation, where stages are discrete classes of say ages or sizes.[1] More recently, the integral projection model (IPM) (Easterling, Ellner, and Dixon 2000; Ellner and Rees 2006, 2007; Rees and Ellner 2009) has emerged as an alternative to matrix projection models when investigating continuous traits, e.g., size, age, mass, leaf length. These models use individual level data to estimate demographic functions, i.e., parametric models for demographic processes specified in the form of vital rates such as growth, maturation, survival, birth, and fertility which are incorporated into a *redistribution kernel*. The term "projection" for these estimated demographic models refers to iterative projection of this kernel to steady state in order to attempt to answer questions regarding *what would happen*. That is, assuming that the kernel yields a steady state, such forward propagation informs about *population statistics*, e.g., analysis associated with the long term population growth rate under a time-invariant projection matrix or kernel. Fitting these models has so far been done only with individual-level transition data which are used to estimate the demographic functions that comprise the *kernel* of the IPM specification. Projection proceeds given the estimated kernel.

The primary perspective of this paper is to argue that such an approach introduces an inherent mismatch in scales. The redistribution kernel in the IPM proposes a mechanistic description of population level redistribution. A kernel of the same functional form, fitted to data at the individual level, would provide a mechanistic model for individual-level processes. Resulting parameter estimates and the associated estimated kernel are at the wrong scale and do not allow population-level interpretation. Expressed in different terms, an individual level model describes the transition of an individual of size x at time t to size y at time $t+1$. It is a conditional specification. As we shall see, an IPM essentially takes a distribution of sizes of individuals at time t to a distribution of sizes of individuals at time $t+1$. Potential confusion arises because in both cases a kernel function is introduced, i.e., a function of x and y. If fitted at the individual level, it describes individual transitions; if fitted at the population level, it reweights a distribution at time t to get a value associated

[1] In the literature the terminology 'matrix population model' is also widely used, e.g., Caswell (2001). We use the notation MPM regardless of the choice for P.

with size y at time $t + 1$. Why should a reweighting based upon individual level data provide an appropriate redistribution at the population level? It seems that we should fit a different model at the latter level. We are dealing with a version of the familiar ecological fallacy (Wakefield 2009).

As a concrete process-level illustration, consider a density dependent kernel fitted to growth and survival of individuals. Such a kernel would say that individuals with few neighbors grow rapidly and vice versa—there is strong density dependence at the individual scale. This heterogeneity in local competitive environments is the basis for most models of coexistence, such as 'successional niche' and 'competition-colonization'. Used at the population scale, the density dependence embedded in the kernel operates on the aggregate density. But this is not the scale where density dependence locally operates, and it cannot capture the role of competition. The fact that individuals experience different local environments—the variation used to fit the kernel—is lost. Instead of capturing this effect, we impose an unrealistic one: when the average density of the plot increases, all individuals respond according to the same kernel. This aggregation problem arises frequently in ecology, hiding or changing relationships between variables (Clark et al. 2011b). Section 2.2 below further elaborates this scaling issue. Offering a remedy is the primary contribution of this paper. In brief, we view the observed size distribution at a given time as a point pattern over a bounded interval. We build a three-stage hierarchical model to infer about the dynamic intensities used to explain the observed point patterns. This model is driven by a *latent*, equivalently, unknown deterministic IPM and we introduce uncertainty by having the operating IPM vary around this deterministic specification. Further uncertainty arises in the realization of the point pattern given the operating IPM. We argue that such dynamic modeling, *optimized* by fitting data observed over time, will more accurately reveal how intensities change over time. That is, there is no mechanism in current IPM model fitting to align projected trait distributions with trait distributions observed over time; the trait distribution obtained from fitting the individual-level IPM does not capture the evolution of the trait distribution from one time point to the next. We show below that forecasts with the latter can drift consequentially compared with those of the former.

We fit the three-stage model within a Bayesian framework. By conditioning on the observed point patterns of sizes, we anchor the parameters of the IPM to the trait distributions observed over time, ensuring that the estimated model is "scaled" to the observed data. Under the Bayesian framework, full inference is available for all model features—all model parameters, all demographic models that comprise the IPM, all predictions of interest under the IPM. If we assume a kernel that is constant over time, then with iteration, we can implement projection, analogous to what is currently done. However, we may prefer time dependent kernels, introducing time-varying environmental information to better explain the observed trait distributions.

The fitting only requires *marginal* point patterns. Apart from a potential data collection advantage, we assert that this is the appropriate way to fit IPM's. That is, our modeling is not a default in the absence of individual level data. Rather, since IPM's "redistribute" intensities (not individual transitions) from one time point to the next, we should be fitting them at aggregated scale, interpreting the kernels at the population scale. If we have individual transition data, we recommend fitting an individual level model, e.g., a dynamic

model such as in Clark et al. (2010a) and references therein. An advantage to working with marginal point patterns is that we need not have observations for all of the years for which we are fitting the IPM. The years for which data are available are incorporated into the likelihood to inform about the parameters in the IPM kernel, but the IPM specification itself can drive the dynamics for years when there are no observations. An implicit assumption is that, within plot, we sample all individuals for the years that we sample; we need to see the complete point pattern for the plot. This is because the number of individuals is assumed random when we model the point pattern given the intensity. Our inference will be at the plot scale. However, since intensity cumulates, we can extrapolate to populations at larger scales. Furthermore, it may be possible to accommodate partial sampling by adjusting intensities for sampling effort (see, e.g., Chakraborty et al. 2011).

Not surprisingly, model fitting under our setting is challenging; an explosion of integrations (or sums) arises. However, we show that, in certain cases, moving to the spectral domain facilitates fitting. Still, fitting is slow, so we offer a *approximate* pseudo-IPM fitting approach that runs more quickly and offers the possibility of more flexible choices for the time-dependent kernels. We develop the foregoing agenda over the next five sections.

The remainder of this introduction provides a brief literature review. MPMs are the most widely used approach to specify population structure. Given the state of the population at time t as a vector of binned cell counts, $\mathbf{n}(t)$, one multiplies it by a *population projection matrix*, $\mathbf{A}$ to yield the state of the population at time $t+1$;

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t). \tag{1.1}$$

If the population is classified into age-classes then the projection matrix is the well-known Leslie matrix. In its simplest form, the projection matrix is assumed to be constant, yielding a linear time-invariant system of difference equations to describe the evolution of the population. A more general, time-varying difference equation version is obtained when one allows the projection matrix to vary because of external factors independent of the state of the population. Further sophistication is achieved when one makes the projection matrix dependent on the current state of population itself, $\mathbf{n}(t+1) = \mathbf{A}_n\mathbf{n}(t)$. This leads to a non-linear model termed a density-dependent MPM. Tuljapurkar and Caswell (1997) and Caswell (2001) discuss the features of all these MPMs in detail. Caswell (2008) provides a general approach to examine the change in responses of a non-linear matrix population models to changes in its parameters.

In any event, the MPMs specify the demographic trait in the form of categorical classifications, so-called stages. In the case of a continuous trait, such as those we have mentioned above and, in particular, for sizes, the classes are ordinal with definition being somewhat arbitrary. Easterling, Ellner, and Dixon (2000) and Ellner and Rees (2006) discuss this issue in detail, noting that the IPM is proposed to remove the categorization required under the MPM approach. Ellner and Rees (2009) remark that "if a trait varies continuously in the study population, then model it that way using an IPM rather than a conventional matrix model."

Using notation from Easterling, Ellner, and Dixon (2000), we denote the continuous trait by x and the distribution of the trait at time t by $n(x,t)$. The projection matrix

is then replaced by a projection kernel $K(y;x)$ that describes how population is redistributed/projected from state x at time t to state y at time $t+1$. With L and U denoting, respectively, lower and upper limits for the value of the trait, the resulting model takes the form

$$n(y,t+1) = \int_L^U K(y;x)n(x,t)\,dx, \tag{1.2}$$

formally specifying an Integral Projection Model (IPM) and yielding a continuous analog of the MPM. Basic theory for a general class of deterministic IPMs, allowing complex demography, is presented in Ellner and Rees (2006). They take K as $K(y,x;\boldsymbol{\theta})$ which is written in terms of parametric specifications for the vital rates. Recent work (Dahlgren, Garcia, and Ehrlén 2011) models these rates using restricted cubic splines. In Ellner and Rees (2007) extension is presented to allow time-varying K's in the form $K(y,x;\boldsymbol{\theta}(t))$ with $\boldsymbol{\theta}(t)$ a random parameter vector. Asymptotic behavior under stationary ergodic $\boldsymbol{\theta}(t)$ is developed. Rees and Ellner (2009) offer a more applied development, supplying fixed and random effects models and showing explicitly that customary growth rate, sensitivity, and elasticity analysis can be obtained with IPMs. In very recent work, Dalgleish et al. (2011) incorporate time varying climate information in the projection kernel and Adler, Ellner, and Levine (2010) develop a multispecies IPM incorporating competition, density dependence, and spatial structure in their individual level model.

IPMs, whether deterministic or stochastic, are an example of a larger class of models usually referred to as integro-difference equations (IDEs). These models have the structure, $W_{t+1}(y) = \int_D h(y,x,;\theta)g(W_t(x))\,dx$. W_t is a stochastic process, and h is a kernel (often a Gaussian) that describes how the process realization at time t is redistributed to a process realization at time $t+1$. The dynamics are controlled by the properties of the redistribution kernel. Typically, $g(z)$ is taken to be ρz in the spirit of stationary time series. Richer choices for h allow the kernel to be local, perhaps θ to be local, and perhaps for the kernel to change over time. In general, although the IDE equations are quite powerful for describing complicated ecological processes (Kot, Lewis, and van den Driessche 1996), with the exception of Wikle and colleagues, e.g., Wikle (2002), Xu, Wikle, and Fox (2005), Hooten et al. (2007) and Wikle and Hooten (2010) they have not often been applied to data in a formal statistical framework. A key distinction is that the kernel h in the IDE setting arises from time discretization of a forward-in-time differential equation while the IPM employs an h that looks back in time to explain how we got to where we are.

As specified, the IPM above is deterministic, raising the question of where and how to insert uncertainty. Within the Bayesian framework, a natural choice is to make the parameters random. To allow for time-varying redistribution kernels, at least two approaches emerge. The first assumes that a vector of parameters is randomly chosen at each time point so that K_t takes the form $K(y,x;\boldsymbol{\theta}(t))$. As noted above, this strategy is employed in, e.g., Rees and Ellner (2009) where, under parametric modeling for $K(y,x;\boldsymbol{\theta})$, the posterior for $\boldsymbol{\theta}$ provides draws for $\boldsymbol{\theta}(t)$. These draws may be interpreted as providing temporal random effects rather than parameter uncertainty. A second approach, which we adopt below (Section 4), is to assume that K is specified as a fixed parametric function but involving time-varying covariates. That is, K_t takes the form $K(y,x;\mathbf{z}_t,\boldsymbol{\theta})$ where $\mathbf{z}_t$ is a vector of

levels of a suitable set of environmental covariates at time t. This approach can be viewed as regression modeling for K.

A broader concern involves uncertainty associated with the form of K itself, hence with the form induced for the $n(y,t)$'s using (1.2). Working with point patterns for the traits, we view the outcome of the IPM as a sequence of intensities and so, at time t, we replace $n(y,t)$ with an intensity function, $\gamma_t(y)$. We view the "operating" intensity, $\lambda_t(y)$ (i.e., the intensity that drives the observed point patterns) as varying around $\gamma_t(y)$ in a fashion described below (Section 4). In other words, insisting that the IPM model is *correct* (even with "best" parameter estimates) is too restrictive. It is easier and more direct to specify uncertainty through the γ's than through the K's. In any event, propagating n's (for us, γ's) through K_t's in (1.2) will not yield explicit forms. In fact, starting at time 0, at time t we have a t dimensional integration for γ_t, requiring approximation to enable tractable computation.

Finally, the format of the paper is as follows. Section 2 fills in the requisite background on dynamic population models. In Section 3 we discuss types of data and, in particular, the Duke Forest data used as an example. Section 4 spells out the modeling details. Section 5 discusses computational approximation with simulated examples and the Duke Forest data analysis in Section 6. Section 7 concludes with a summary and future directions.

2. THE MPM AND IPM APPROACH IN ECOLOGICAL DEMOGRAPHY

MPM's and IPM's have become the techniques of choice for ecological demography. These models are specified with two indices, one for time, the other for trait level. There can be continuity or discreteness in time as well as continuity or discreteness in the trait space. With discrete time and discrete (categorical) trait space we have a MPM; with continuous time and discrete trait space an ordinary differential equation (ODE); with discrete time and continuous trait space an integro-difference equation (IDE); and with continuous time and continuous trait space a partial differential equation (PDE). As noted in the Introduction, IPM models are a subset of the IDE models and, thus, since an IDE can be developed through discretization of a PDE, similarly it can be the case for an IPM. However, an arguably useful distinction between IPMs and IDEs is the forward vs. backward perspective. IDE's look forward using structured kernels to describe where the process is going. That is, under time-discretization of a PDE, the forward dynamics induce a redistribution kernel to yield an IDE. On the other hand, IPMs look backward, using structured, mechanistic, process-driven kernels to describe how the process got to where it is. With discretization of the trait scale, this distinction is preserved. For the IDE, the kernel becomes a *propagator* matrix; for the IPM we obtain a MPM.

2.1. MATRIX PROJECTION MODELS

Matrix projection models are the most widely used demographic approach for structured biological populations. The population vector $\mathbf{n}(t)$ lists the numbers of individuals in a finite set of categories at time t. Again, the projected population in these categories at

time $t+1$ is given by $\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t)$. The $\mathbf{A}_{ij}$ give the average per-capita contribution from individuals in category j at time t to category i at time $t+1$, either by survival, growth, or reproduction. In fact, typically, $\mathbf{A}$ is written in the form $\mathbf{A} = \mathbf{T} + \mathbf{F}$ with $\mathbf{T}$ describing transition (survival and growth) and $\mathbf{F}$ describing reproduction (fecundity). Under the model, the stationary behavior of this matrix projection equation is obtained in terms of the eigenvalues (Λ_i) and eigenvectors ($\mathbf{w}_i$) of the projection matrix $\mathbf{A}$. The long-term behavior of $\mathbf{n}(t)$, i.e., the ergodic properties of population growth, is determined by the dominant eigenvalue, $\max(\Lambda_i)$ and associated right eigenvector. Generalization of this simple MPM to admit variability in survival and fecundity rates over time leads to the stochastic MPM given by $\mathbf{n}(t+1) = \mathbf{A}(t)\mathbf{n}(t)$ where $\{\mathbf{A}(t),\ t = 0, 1, 2, \ldots\}$ is a stochastic sequence of non-negative matrices. Under suitable assumptions, this stochastic MPM still enables a dominant eigenvalue that is almost surely constant (Tuljapurkar 1990). Further eigenanalysis of the projection matrix yields a set of population statistics, viz., population growth rate, damping ratio, reproductive value and so on (Caswell 2001). Evidently, in analyzing populations using projection with MPMs, estimation of the eigenvalues and eigenvectors are of critical importance.

In general, when the model is density dependent, i.e., $\mathbf{n}(t+1) = \mathbf{A}_n\mathbf{n}(t)$, the resulting behavior of the matrix equation cannot be written in terms of eigenvalues and eigenvectors (Caswell 2001, p. 504). Under a simplifying density dependence assumption, one can obtain the dominant eigenvalue and eigenvector of $\mathbf{A}_n$ using an equilibrium condition. But when the dependence is complex, an equilibrium solution is extremely difficult to obtain.

If, however, one has a time series of population vectors $\mathbf{n}(t)$, without the individuals being identifiable, a maximum likelihood approach can be used to estimate the elements of $\mathbf{A}$. Dennis et al. (1995, 1997) achieved this by introducing a stochastic component in the matrix equation as follows:

$$\mathbf{n}(t+1) = \exp\big(\mathbf{D}(t)\big)\mathbf{A}_n\mathbf{n}(t),$$

where $\mathbf{D}(t) = \operatorname{diag}(d_1(t), \ldots, d_s(t))$ are assumed to be drawn from multivariate normal distribution with mean 0 and covariance Σ. The resulting log-normal model yields a likelihood function which can optimized to obtain the maximum likelihood estimates of the model parameters. We see that $\mathbf{D}(t)$ imposes process after demography.

2.2. THE IPM AND ASSOCIATED PROPERTIES

Here, we investigate the behavior of an IPM as a deterministic specification. In fact, the form in (1.2) has been referred to as a population-level *mean field model*, tracking expected number of individuals and their associated expected state distribution (Ellner and Rees 2007). The kernel $K(y, x; \boldsymbol{\theta})$ is the IPM analog of the projection matrix $\mathbf{A}$ in MPM. We rewrite (1.2) in terms of intensities, subscripted by time, i.e.,

$$\gamma_{t+1}(y) = \int_L^U K(y; x)\gamma_t(x)\,dx. \tag{2.1}$$

To give a population level interpretation to (2.1), it may be easiest to think in terms of *intensity elements*. That is, $\gamma_{t+1}(y)\,dy = \int_X K(y; x)\,dy\,\gamma_t(x)\,dx$. But then, we see that

$K(y;x)\,dy\,\gamma_t(x)\,dx$ is the expected number of individuals in size interval $(y, y+dy)$ at time $t+1$ from all individuals in size interval $(x, x+dx)$ at time t.

To study the properties of a deterministic specification, we suppress parameters in (2.1). Let $\gamma_{t,\cdot} = \int_L^U \gamma_t(x)\,dx$, i.e., $\gamma_{t,\cdot}$ is the expected number of individuals at time t. Integrating (2.1) over y from L to U yields $\gamma_{t+1,\cdot} = \int_L^U K(\cdot,x)\gamma_t(x)\,dx$, where $K(\cdot,x) = \int_L^U K(y;x)\,dy$; $\gamma_{t+1,\cdot}$ can be compared with $\gamma_{t,\cdot}$.[2] Integrating (2.1) over $B \subset [L,U]$ yields $\gamma_{t+1}(B) = \int_L^U K(B,x)\gamma_t(x)\,dx$.

The eigenvalue theory for the IPM can be directly connected to that for the MPM by viewing $K(y;x)$ as a linear operator, i.e., $Kh \equiv \int_L^U K(y;x)h(x)\,dx$. Then, if Λ is the largest eigenvalue associated with K and $w(x)$ is the associated right eigenfunction,[3] $\int_L^U K(y;x)w(x)\,dx = \Lambda w(y)$ showing that, at steady state, Λ is the growth rate and $w(x)$ (normalized) is the steady state size distribution. As a result, $K^t w = \Lambda^t w$. However, for a given t and an arbitrary initial size distribution $\gamma_0(x)$, $K^t\gamma_0$ need not be close to $\Lambda^t\gamma_0$.

A common mechanistic specification for K, viewing the process at the individual-level, takes the form

$$K(y;x) = q(x)f(y|x) + \Delta(x)g(y). \tag{2.2}$$

These terms are interpreted as follows. The first pair capture survival and growth; $q(x)$ is the probability of survival of an individual of size x and $f(y|x)$ is the conditional probability density of size y next year, given current size x of an individual. The second pair capture fecundity or recruitment (in our Duke forest example below); $\Delta(x)$ is the number of recruits associated with individuals of size x and $g(y)$ is the probability density of sizes for the new recruits. In the literature, these components have been modeled at the individual level. For instance, individual level data are used to fit a logistic regression for q and a Poisson regression with log link for Δ. Also, a Gaussian regression using individual transitions is used to fit f and an exponential function, using individual seed production, is used to fit g.

Now, we can explicitly clarify the distinction between current IPM work and our approach. In the sequel, we still use the form in (2.2) but, as per the clarification above using intensity elements, we interpret q, f, Δ, and g so that the resulting K is interpreted as redistributing the intensity at year t to a new intensity at year $t+1$. (See Section 4 below for details.) Again, we think in terms of population transition, not individual transition.

Inserting (2.2) into (2.1) yields

$$\gamma_{t+1}(y) = \int_L^U \big(q(x)f(y|x) + \Delta(x)g(y)\big)\gamma_t(x)\,dx. \tag{2.3}$$

Under (2.3), we find that $\gamma_{t+1,\cdot} = \int_L^U K(\cdot,x)\gamma_t(x)\,dx = \int_L^U (q(x)+\Delta(x))\gamma_t(x)\,dx$. If $q(x) = q$ and $\Delta(x) = \Delta$, i.e., constant rates over $[L,U]$, we obtain $\gamma_{t+1,\cdot} = (q+\Delta)\gamma_{t,\cdot}$ so

[2]Note that L can be 0 but we must have $U < \infty$ in order to ensure that the γ_t's are valid intensities. This should be a mild practical constraint.

[3]Perron–Frobenius theory tells us that, at this Λ, $w(x) \geq 0\ \forall x$.

$q + \Delta$, greater than, less than or equal to one determines whether the expected population size grows, decreases or remains constant. In practice, constant values are not plausible; we will want forms $q_t(x)$ and $\Delta_t(x)$ where the dependence on t arises through the introduction of covariate information at time t as well as, perhaps, through introduction of density dependence.

Sometimes normalization is introduced into the IPM. One possibility is to replace $K(y; x)$ in (2.1) with say $K(y; x)/K(.; x)$, a normalized version. This is too restrictive; it removes the interpretation of $\gamma_t(x)$ as an intensity since it imposes $\gamma_{t,.}$ constant over t. Normalizing $\gamma_t(x)$ to the density $\tilde{\gamma}_t(x) = \gamma_t(x)/\gamma_{t,.}$ is also unattractive since it now normalizes the resulting $\gamma_{t+1}(y)$ by $\gamma_{t,.}$ rather than by $\gamma_{t+1,.}$. In this regard, one might consider viewing $K(y; x)$ as a conditional intensity on $[L, U]$ for y at a given x, whence $K(y; x)\gamma_t(x)$ would become a joint intensity over $[L, U] \times [L, U]$. This is not appropriate for our setting. A joint intensity produces (x, y) pairs; we do not impose $\gamma_{t+1,.} = \gamma_{t,.}$.

Note that if $K(y; x)$ is of the form $K(y - x)$ and, *formally*, we allow the limits of the integration to expand to $(-\infty, \infty)$, then $\gamma_{t+1}(y)$ arises as a convolution of K and γ_t. This reveals that, if we start with γ_0, we will obtain a t-fold integral in order to arrive at γ_t. Computation, using discretization, will result in a t-fold sum and will become infeasible. However, we can take advantage of the convolution representation to work in the Fourier domain. That is, with tilde denoting the associated Fourier transform, we have

$$\tilde{\gamma}_{t+1}(u) = \int_{-\infty}^{\infty} e^{-iyu}\gamma_{t+1}(y)\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-iyu}K(y - x)\gamma_t(x)\,dx\,dy = \tilde{K}(u)\tilde{\gamma}_t(u)$$

and, more generally, $\tilde{\gamma}_t(u) = (\tilde{K}(u))^t\tilde{\gamma}_0(u)$. Hence, we can replace multiple integration/summation with multiplication and then implement a one-dimensional inverse Fourier transformation to recover $\gamma_t(y)$. Explicit details under specific forms for K (and, in fact, extended to $K_t(y - x)$) with γ_0 are discussed in Section 5.2 below.

2.3. DENSITY DEPENDENT IPMS

Density dependence is introduced into an IPM to capture the effect of competition on population growth and structure. At time t it revises $K(y; x)$ to $K(y; x, \gamma_{t,.})$. In (2.2), we expect survival probability and expected recruitment at time t to be affected by population size at time t (perhaps f as well). Suppose $q_t(x) = q(\gamma_{t,.})$ and $\Delta_t(x) = \Delta(\gamma_{t,.})$. From above, we have $\gamma_{t+1,.} = (q(\gamma_{t,.}) + \Delta(\gamma_{t,.}))\gamma_{t,.}$, revealing how, with suitable choices of $q(\cdot)$ and $\Delta(\cdot)$, expected population size at time t will affect expected population size at time $t + 1$. If $q(\gamma_{t,.}) + \Delta(\gamma_{t,.}) > 1$, we expect growth in population size from year t to year $t + 1$, the opposite if < 1.

Current understanding of demographic rates, population growth, and density dependence provides some guidance on functional forms. Both $q(\gamma)$ and $\Delta(\gamma)$ should be strictly decreasing (see Section 4). As $\gamma \to \infty$ both should tend to 0. Potentially, as $\gamma \to 0$, q could tend to 1 and Δ to $\Delta_0 > 0$. For tree populations in the example below, $\frac{\gamma_{t+1,.}}{\gamma_{t,.}} = q(\gamma_{t.}) + \Delta(\gamma_{t,.}) \in [1 - \epsilon, 1 + \epsilon]$ where ϵ might be 0.1. That is, barring catastrophes such as hurricanes or logging, a change in forest size of more than 10 % over a year does not occur. Furthermore, survival probabilities are high and, hence Δ is small. This

implies that, practically, we need not be concerned about behavior of our models for $q(\cdot)$ and $\Delta(\cdot)$ at extremes of the γ scale.

Two common forms for rates in the literature are $1/(1+c\gamma)$ (Beverton–Holt) and $e^{-c\gamma}$ (Ricker) with $c>0$ (see, e.g., Caswell 2001, Chapter 16), yielding, in the latter case, $q(\gamma)=e^{-c_q\gamma}$ and $\Delta(\gamma)=e^{-c_\Delta\gamma}$. We note that, in this context, Caswell (2001) has suggested an overall scaling to the kernel which, in our notation, would take the form $\gamma_{t+1}(y)=\int K(y;x)b(x,\gamma_{t,\cdot})\gamma_t(x)\,dx$, viewing b as a per-capita growth rate. A more general Ricker growth rate specification is associated with the Ricker curve (1954). It takes the form $\gamma e^{r(1-\frac{\gamma}{k})}$ with r interpreted as the growth rate and k the carrying capacity.

Survival probability and recruitment in year t would be expected to depend on environmental variables $\mathbf{z}_t$ along with population size. Hence, we can extend q to $q(\gamma_t,\mathbf{z}_t)$, similarly for Δ. For q, consider a logit, i.e.

$$q(\gamma,\mathbf{z})=\frac{\exp(c_0+c_1\gamma+\beta_q\mathbf{z})}{1+\exp(c_0+c_1\gamma+\beta_q\mathbf{z})}.$$

At $\beta=0$ and $c_1<0$, we have an alternative two parameter choice to the Ricker curve, which also has the benefit of convenient regression modeling. Note that if $\beta=0$, as $\gamma\to 0$, q does not approach 1, but can be arbitrarily close according to c_0. But, again, within the range of the $\gamma_{t,\cdot}$'s in our data, this will not be a concern. So, below, we work with logit forms and with $C_0=\exp(c_0)$ and $C_1=-c_1$. For $\Delta(\gamma,z)$, we can use the Ricker form directly; on the log scale, we take $\log\Delta(\gamma,z)=\delta_0+\delta_1\gamma+\beta_\Delta\mathbf{z}$.

Lastly, we note that Ellner and Rees (2006) generalized the deterministic IPM by assuming a density dependent kernel $K(y,x,\mathbf{N}_t)=P(y,x)+h(\mathbf{N}_t)\Delta(x)g(y)$ where $\mathbf{N}_t$ is a weighted total population size given by $\mathbf{N}_t=\int_x W(x)n(x,t)\,dx$ for some weighting function $W\geq 0$. Under this formulation of density dependence, they obtained the population growth rate and net reproductive rate.

3. DATA TYPES AND THE DUKE FOREST DATA

As discussed in, e.g., Caswell (2001), with MPMs, demographic data are customarily in one of two forms—most commonly at the individual level, i.e., individual level dynamics over time, or, occasionally, in the form of a time series of population vectors. In the first case, the individual level observations are used to fit parametric demographic models for vital rates. Once the parameters are estimated, they can be inserted into the MPM in (1.1) to project the population status over time. With data of the second type, we observe a sequence of population vectors $\mathbf{n}(t_1),\mathbf{n}(t_2),\ldots$ over time without distinguishing the individuals. As in Section 2.1, model fitting in this setting is discussed by Dennis et al. (1995, 1997) who consider time as stage duration and introduce stochastic noise in the deterministic difference equations that describe the population dynamics. Then, they appeal to the theory of non-linear multivariate time-series methodology to obtain the maximum likelihood estimates of the model parameters. This likelihood-based methodology allows one to statistically test the adequacy of the deterministically specified model for population evolution, conditional on the available data. Also in this spirit, though in the context of PDEs, is work of Banks et al. (1987, 1991).

Like the MPM, fitting IPM's is typically done with individual data. Data of the second type would consist of a time series of point patterns for the trait distribution, e.g., size, associated with a specified geographic region. So far as we are aware, only data of the first type have been employed for fitting IPM's. As mentioned in the introduction, here we propose to work only with data of the second type.[4] With individual level data, perhaps the most direct modeling strategy would be through a dynamic model with individual level random effects as in, e.g., Clark et al. (2010a). The state-space framework provides inference on individual variation in terms of population parameters, while being anchored directly by observations at the same scale. Afterward, desired population-level summaries can be created. In this regard, the IPM, viewed as a mean field model, is not providing average or expected transition behavior associated with a state space specification.

Hence, we see the role and advantage of IPM's for large studies where tracking of individuals is not feasible. For instance, we could not hope to track individual trees in multiple forests on an annual basis over a span of many decades. Collecting marginal point patterns at the scale of plots, without transition information on individuals, is more realistic. Hence, we need to be able to fit IPMs with data of the second type. In particular, in the application that follows we use diameter distributions of *Liriodendron tulipifera* (tulip poplar) from the Duke Forest, North Carolina. Data collection is detailed in Clark et al. (2010b). A sample of individual trees are marked and remeasured in this longitudinal study, but, for the IPM we fit, we use the entire annual inventories, yielding annual (marginal) observation of sizes. This form would arise from periodic surveys where individuals are not marked.

4. MODEL DETAILS

Since the IPM models described in (2.2) approximate population demography, it is not realistic to assume that they provide the true intensities that are driving the point patterns. So, we introduce uncertainty, which can be accomplished in several ways. One might attempt to do this within the kernel specification. However, even with the mechanistic forms in (2.2), this will prove computationally infeasible. With regard to the intensities, we cannot propose $\lambda_t(x) = \gamma_t(x) + \epsilon_t(x)$ with say, $\epsilon_t(x)$ a stationary mean 0 Gaussian process over $[L, U]$ with covariance function $\sigma_\epsilon^2 \rho(\cdot, \phi)$ because this could allow $\lambda_t(x) < 0$. So, as is common with a stochastic process of positive random variables, we introduce the Gaussian process on the log scale, i.e., $\lambda_t(x) = \gamma_t(x) e^{\epsilon_t(x)}$, where ϵ_t is the mean 0 Gaussian process above. An alternative specification would be a Gaussian process truncated at 0 but this would be much more computationally demanding.

Furthermore, we do not apply the $K_t(y, x)$ to $\lambda_t(x)$. Rather, we allow the IPM to provide dynamics in a deterministic fashion for the γ_t's and view the λ_t's as varying around their respective γ_t's. A potential benefit is the possibility to take advantage of the recursive relationship mentioned in Section 2, which may be available under the deterministic IPM.

[4] A simple statistical analogue would be paired vs. nonpaired comparisons in a two-sample problem.

It is also facilitates a pseudo-IPM approximation, as discussed in Section 5. As a result, the λ_t's are conditionally independent given $\{\gamma_t\}$. At the highest level, we assume the point patterns, the $\mathbf{x}_t$'s, are conditionally independent given $\{\lambda_t\}$ and that each follows the usual density for a nonhomogeneous Poisson process given its intensity. That is, the likelihood associated with the time point t is given by

$$[\mathbf{x}_t|\lambda_t] \propto \left[\exp\left(-\int_L^U \lambda_t(x)\,dx\right)\prod_{i=1}^{n_t}\lambda_t(x_{ti})\right]. \tag{4.1}$$

So, the observed sizes are conditionally independent given $\lambda_t(x)$ but marginally dependent due to the log Gaussian process model for $\lambda_t(x)$. This convenience in specification demonstrates the benefit of the multi-level modeling.

We initiate our model with γ_0, a kernel intensity estimate (Diggle 2003). Hence, the full posterior is proportional to

$$\begin{aligned}&\prod_{t=1}^{T}\left[\mathbf{x}_t\left|\lambda_t(x), x\in[L,U]\right.\right]\\ &\quad\times\prod_{t=2}^{T}\left[\lambda_t(x), x\in[L,U]\left|\gamma_t(x), x\in[L,U], \sigma^2, \phi\right.\right]\\ &\quad\times\left[\left\{\gamma_t(\boldsymbol{\theta},\gamma_0), t=1,\ldots,T\right\}\right]\left[\sigma^2\right][\phi][\boldsymbol{\theta}].\end{aligned} \tag{4.2}$$

In (4.2), the bracketed term involving $\{\gamma_t\}$ is a degenerate distribution. It is only employed to denote the deterministic functional specification for the γ_t's given the IPM and $\boldsymbol{\theta}$. In fitting the model, it is computationally efficient to marginalize over the λ_t's using Laplace approximation, as described below. Typically, our interest will be in the process, i.e., in $K_t(y,x;\boldsymbol{\theta})$, particularly its components and associated parameter estimates, as well as the resultant $\gamma_t(x)$'s. Of course, we can always obtain the λ_t's after the fact by appropriate composition sampling (Banerjee, Carlin, and Gelfand 2004).

In specifying the redistribution kernel K we confine ourselves to a parametric form. Following Section 2, we assume K to be comprised of growth and recruitment and, with density dependence, we write

$$K_t(y,x;\mathbf{z}_t,\boldsymbol{\theta},\gamma_{t,\cdot}) = G_t(y,x;\mathbf{z}_t,\boldsymbol{\theta},\gamma_{t,\cdot}) + R_t(y;\mathbf{z}_t,\boldsymbol{\theta},\gamma_{t,\cdot}), \tag{4.3}$$

where the growth term $G_t(.)$ is further decomposed as

$$G_t(y|x;\mathbf{z}_t,\boldsymbol{\theta},\gamma_{t,\cdot}) = q(x,\gamma_{t,\cdot},\mathbf{z}_t)\,f_t(y-x;\mathbf{z}_t,\boldsymbol{\theta}).$$

Again, returning to density elements, we interpret $q(x,\gamma_{t,\cdot},\mathbf{z}_t)\,f_t(y-x;\mathbf{z}_t,\boldsymbol{\theta})\,dx\,dy$ as the expected number of individuals in size interval $(y, y+dy)$ at time $t+1$ from survivors in size interval $(x, x+dx)$ at time t. In particular, we assume f_t to be Gaussian density with mean and variances both depending on the covariates $\mathbf{z}_t$. Note that a "translation-invariant" assumption for f_t can be appropriate at the population level though it would almost never be sensible at the individual level. We assume survival probability declines as function of

$\gamma_{t,\cdot}$ due to resource limitation. In its simplest form, following Section 2 and ignoring x and $\mathbf{z}_t$, we would take q to be

$$q(\gamma_{t,\cdot}) = \frac{Q_0 e^{-Q_1\gamma_{t,\cdot}}}{1 + Q_0 e^{-Q_1\gamma_{t,\cdot}}}, \tag{4.4}$$

where Q_0 and Q_1 (both > 0) are parameters that govern the rate of decay of the survival probability.

The recruitment term takes a form similar to the growth term,

$$R_t(y; \mathbf{z}_t, \boldsymbol{\theta}, \gamma_{t,\cdot}) = \Delta(x, \gamma_{t,\cdot}, \mathbf{z}_t) g_t(y; \mathbf{z}_t, \boldsymbol{\theta}).$$

With density elements, analogously, we interpret $\Delta(x, \gamma_{t,\cdot}, \mathbf{z}_t) g_t(y; \mathbf{z}_t, \boldsymbol{\theta})\, dx\, dy$ as the expected number of recruited individuals in size interval $(y, y + dy)$ at time $t + 1$ from individuals in size interval $(x, x + dx)$ at time t. Customarily, the terms on the right side reflect flowering and seed production (see, e.g., Rees and Ellner 2009). However, in our examples, for the L we work with, seeds in year t will not produce trees of size greater than L in year $t + 1$. Hence, the recruitment simply describes the size intensity for arrivals in year $t + 1$. Δ is the expected influx in year $t + 1$ and g_t is a density on $\tilde{y} = y - L$, which is assumed to be a translated exponential with mean depending on the covariates. We assume influx declines with $\gamma_{t,\cdot}$ due to reduced availability of resources. So a simple form for Δ, following Section 2.3, is

$$\log\Delta(\gamma_{t,\cdot}) = \delta_0 - \delta_1\gamma_{t,\cdot}, \tag{4.5}$$

with δ_0 and δ_1 both positive.

Under (4.4) and (4.5), we achieve a stable population if the expected number of individuals remains approximately the same over time implying $q(\gamma_{t,\cdot}) + \Delta(\gamma_{t,\cdot}) \approx 1, \forall t = 1, 2, \ldots$. A population explosion is experienced when $q(\gamma_{t,\cdot}) + \Delta(\gamma_{t,\cdot}) \gg 1$ and similarly extinction can occur if $q(\gamma_{t,\cdot}) + \Delta(\gamma_{t,\cdot}) \ll 1$. Of course, the case $Q_1 = \delta_1 = 0$ is density independence.

To handle asymmetric growth, one can specify an asymmetric kernel, such as a log-Gaussian distribution for f_t. In fact, one can also make the redistribution kernel richer by specifying a semiparametric or nonparametric model for it but at the expense of more difficult computation for model fitting.

Returning to (4.1), the stochastic integral cannot be evaluated in closed form, so we approximate it with a Riemann sum. We divide the interval $[L, U]$ into a fine grid consisting of B cells of equal length with the centers given by x_j^*. We further assume that the intensity is homogeneous within each cell and that the centers, x_j^*, remain fixed over the entire time period under study. The length and cell level intensity for cell b are denoted by l and $\lambda_t(b)$; $b = 1, \ldots, B$, respectively. Then the operational likelihood in (4.2) becomes

$$\prod_{t=1}^{T}\left[\exp\left(-\sum_{b=1}^{B}\lambda_t(b)l\right)\prod_{b=1}^{B}\left[\lambda_t(b)\right]^{n_{tb}}\right], \tag{4.6}$$

where n_{tb} is the number of points in cell b in year t. The choice of the quadrature does not impact parameter estimates as long as the grid is fine enough to justify the assumption

of homogeneous intensity on each grid cell. However, computational complexity increases with increase in the number of cells. Thus, if the computation is rendered infeasible due number of cells, one would coarsen the resolution and apply a more sophisticated quadrature method to evaluate the integral (see, e.g., Davis and Rabinowitz 1984).

As noted above, we assume $f_t(y-x;\mu_t,\sigma_t^2) = \phi(y-x;\mu_t,\sigma_t^2)$ and $g_t(y;\eta_t) = \eta_t e^{-\eta_t(y-L)}$, $y > L$, with the parameters $\mu_t, \sigma_t^2, \eta_t$ incorporating the covariate information $(\mathbf{z}_t)$. We assume $\mu_t \sim N(\mathbf{z}_t\boldsymbol{\beta}_\mu, \sigma_\mu^2)$, $\log(\sigma_t^2) \sim N(\mathbf{z}_t\boldsymbol{\beta}_\sigma, \sigma_\sigma^2)$, and $\log(\eta_t) \sim N(\mathbf{z}_t\boldsymbol{\beta}_\eta, \sigma_\eta^2)$. These specifications presume no intercept in $\mathbf{z}_t$ and thus, can be interpreted as introducing a time-varying intercept. We adopt independent vague $N(0, 10^2)$ prior for all β's. Instead of assigning priors for $\sigma_\mu^2, \sigma_\sigma^2, \sigma_\eta^2$ we fix them arbitrarily at 10^2 in the present study. Preliminary investigation in fitting this multi-level hierarchical model suggests that posteriors for these variance parameters have very large standard errors. The data carry little information about them so we recommend fixing them at some large values. Under the assumption of the exponential correlation structure, σ_ϵ^2 and ϕ cannot be jointly consistently estimated (Zhang 2004). Since we are more interested in learning about the uncertainty in how λ varies around γ, we propose a weak prior for σ_ϵ^2 and fix the decay parameter ϕ, setting it to correspond to one-third of the range we adopt for the size distribution. We outline a procedure to develop prior information about σ_ϵ^2 in Section 6.1.1.

For the forms in (4.4) and (4.5), imposing priors on $q(\gamma_{t,\cdot})$ and $\Delta(\gamma_{t,\cdot})$ requires specifying priors on Q_0, Q_1, δ_0 and δ_1, respectively. We interpret $\frac{Q_0}{1+Q_0}$ as the survival probability when the population size tends to 0 and $\frac{\delta_0}{1+\delta_0}$ as the replacement rate when the population size tends to 0. By analogy with the Ricker curve (Section 2.3), we can roughly interpret Q_1 to be the global survival probability of the species and δ_1 to be the average rate of influx shown by that species. Q_0, Q_1, δ_0, and δ_1 are not well identified. In fact, intuitively, from Section 2, the sum $q(.) + \Delta(.)$ is well identified but not its components. In Section 6.1.2 we discuss an estimation strategy to handle q and Δ using knowledge of the ecological processes driving the survival and influx for the population.

5. MODEL FITTING AND INFERENCE

As we have noted, the model described in (4.2) is computationally demanding to fit. The computational challenge arises because (2.3) does not have a closed form solution; we need to resort to numerical integration to create the sequence of $\{\gamma_t(x)\}$. From Section 2, the dimension of the numerical integrations increases as the number of time epochs increases and consequently will result in an explosion of terms as we sum. An MCMC scheme will be computationally prohibitive because we will have to perform these integrations iteration by iteration. First, in Section 5.1, we consider a simpler form for K where we can perform exact IPM updating. Then, in Section 5.2 we propose an approximate "pseudo" IPM approach, using adjacent pairs of years, that allows us to handle more general K.

5.1. COMPUTATION FOR A DENSITY INDEPENDENT MODEL

Here, we consider the special case where, in K, the conditional density f, takes the form $f(y-x;\mathbf{z}_t,\boldsymbol{\theta})$ and the q_t and Δ_t are time dependent, e.g., on $\mathbf{z}_t$ but not size or

density dependent.[5] Then, the update can be written as

$$\gamma_{t+1}(y) = \int_D \left[q_t f_t(y-x;\mathbf{z}_t,\boldsymbol{\theta}) + \Delta_t g_t(y;\mathbf{z}_t,\boldsymbol{\theta})\right]\gamma_t(x)\,dx. \tag{5.1}$$

We assume $\gamma_0(x) = \tilde{\gamma}_{0,.}\phi(x;\mu^*,\sigma^{*2})$ where $\tilde{\gamma}_{0,.}$ is the initial population size and $\phi(.;\mu^*,\sigma^{*2})$ is the density function of a Normal distribution with mean μ^* and variance σ^{*2}. (Sensitivity to this initial intensity will, of course, disappear as t increases.) We assume, as earlier, that $f_t(y-x;\mathbf{z}_t,\boldsymbol{\theta}) = \phi(y-x;\mu_t,\sigma_t^2)$ and $g_t(y;\mathbf{z}_t,\boldsymbol{\theta}) = \eta_t e^{-\eta_t(y-L)}$, $y > L$. We further assume that $\mu_t, \sigma_t^2, \eta_t, q_t = q(\mathbf{z}_t,\boldsymbol{\theta}), \Delta_t = \Delta(\mathbf{z}_t,\boldsymbol{\theta})$ are all functions of the covariates and constant over x. Then, taking the Fourier transform of $\gamma_{t+1}(y)$ in (5.1) we arrive at the following result.

Result 1: Let $\tilde{g}(u) = \int_{-\infty}^{\infty} e^{iux} g(x)\,dx$ denote the Fourier transform of the function $g(x)$. Under the above assumptions we have

$$\begin{aligned}
\tilde{\gamma}_t(u) = &\prod_{l=1}^{t-1} q_l e^{iu(\sum_{l=1}^{t-1}\mu_l+\mu^*)-\frac{1}{2}u^2(\sum_{l=1}^{t-1}\sigma_l^2+\sigma^{*2})} \\
&+ \prod_{l=2}^{t-1} q_l e^{iu(\sum_{l=2}^{t-1}\mu_l+L)-\frac{1}{2}u^2(\sum_{l=2}^{t-1}\sigma_l^2)}\gamma_{1,.}\Delta_1\eta_1/(\eta_1-iu) \\
&+ \prod_{l=3}^{t-1} q_l e^{iu(\sum_{l=3}^{t-1}\mu_l+L)-\frac{1}{2}u^2(\sum_{l=3}^{t-1}\sigma_l^2)}\gamma_{2,.}\Delta_2\eta_2/(\eta_2-iu) \\
&+ \cdots \\
&+ q_{t-1} e^{iu(\mu_{t-1}+L)-\frac{1}{2}u^2\sigma_{t-1}^2}\gamma_{t-2,.}\Delta_{t-2}\eta_{t-2}/(\eta_{t-2}-iu) \\
&+ \gamma_{t-1,.}\Delta_{t-1} e^{iuL}\eta_{t-1}/(\eta_{t-1}-iu),
\end{aligned} \tag{5.2}$$

where $\gamma_{t,.} = \int_L^U \gamma_t(x)\,dx$. The proof follows by direct calculation.

Using this recursion relation in the spectral domain we can efficiently compute the whole sequence of $\{\tilde{\gamma}_t(u)\}$, alleviating the problem of explosion of integrals mentioned above. We employ numerical integration to invert the Fourier transform and recover $\{\gamma_t(x^*)\}$'s, as described in the Appendix.

We assume the same priors for $\mu_t, \sigma_t^2, \eta_t$, and the β's as in Section 4. Also, we specify priors on the survival probability, $\text{logit}(q_t) \sim N(\mathbf{z}_t\boldsymbol{\beta}_q, \sigma_q^2)$, and expected influx $\text{logit}(\Delta_t) \sim N(\mathbf{z}_t\boldsymbol{\beta}_\Delta, \sigma_\Delta^2)$. Again, these priors can be interpreted as introducing time-dependent intercepts for q_t and Δ_t. As before, we assume the parameters, $[\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_q, \boldsymbol{\beta}_\Delta, \boldsymbol{\beta}_\eta]$, to be mutually independent and impose a vague normal prior (mean 0, variance 100) on them; We fix $\sigma_\mu^2, \sigma_\sigma^2, \sigma_q^2, \sigma_\Delta^2, \sigma_\eta^2$ arbitrarily at 10^2. Since we have already gridded the interval $[L, U]$, we can obtain a frequency distribution of the trait and can thereby evaluate the

[5] Here and below, we use $\boldsymbol{\theta}$ to denote the collection of parameters in (4.2), acknowledging that a suitable subset of them is associated with each of the components of K.

modal value of that trait. In the present situation, we fix $\mu^* = \text{mode}(\mathbf{x}_1)$. Since σ^{*2} controls the spread of the intensity curve associated with the initial population vector, we set it at $\text{var}(\mathbf{x}_1)$.

A further simplification is obtained if we can integrate out λ_t from the joint posterior arising from (4.2) using Laplace approximation. Then, we can draw $\boldsymbol{\theta}$ from the resulting approximate marginal posterior distribution. The following result helps us to obtain this approximate distribution.

Result 2: Under discretization, we replace $\boldsymbol{\lambda}_t = \{\boldsymbol{\lambda}_t(x), x \in [L, U]\}$ with $\boldsymbol{\lambda}_t$, a $B \times 1$ vector, following (4.5), similarly with $\boldsymbol{\gamma}_t$. Then, at a particular time point t, an approximation for the conditional distribution of $\mathbf{x}_t$ given $\boldsymbol{\theta}$, marginalized over λ_t is given by

$$[\mathbf{x}_t|\boldsymbol{\theta}] \approx \left[\frac{|\Sigma_t^*|}{|\Sigma|}\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left((\log \boldsymbol{\gamma}_t)'\Sigma^{-1}(\log \boldsymbol{\gamma}_t) - \boldsymbol{\gamma}_t^{*\prime}\Sigma^{-1}\boldsymbol{\gamma}_t^*\right)\right] \times \exp\left[-\frac{1}{2}\left(\boldsymbol{\eta}_t^* - \boldsymbol{\gamma}_t^*\right)'\Sigma_t^{*-1}\left(\boldsymbol{\eta}_t^* - \boldsymbol{\gamma}_t^*\right) - \sum_{i=1}^{b} e^{\eta_{ti}^*}\right], \tag{5.3}$$

where $\boldsymbol{\gamma}_t^* = \log \boldsymbol{\gamma}_t + \Sigma \times \mathbf{n}_t$ with $\mathbf{n}_t = [n_{t1}, \ldots, n_{tB}]'$. $\Sigma_t^{*-1} = [\Sigma^{-1} + \Sigma_2(\boldsymbol{\gamma}_t^*)]$ and $\boldsymbol{\eta}_t^* = \boldsymbol{\gamma}_t^* - [\Sigma^{-1} + \Sigma_2(\boldsymbol{\gamma}_t^*)]^{-1} J_2(\boldsymbol{\gamma}_t^*)'$. The expressions for $\Sigma_2(\boldsymbol{\gamma}_t^*)$ and $J_2(\boldsymbol{\gamma}_t^*)$ are given by

$$\Sigma_2\left(\boldsymbol{\gamma}_t^*\right) = l \times \text{diag}\left(le^{l\gamma_{t1}^*}, \ldots, le^{l\gamma_{tB}^*}\right),$$
$$J_2\left(\boldsymbol{\gamma}_t^*\right) = l \times \left[e^{l\gamma_{t1}^*}, \ldots, e^{l\gamma_{tB}^*}\right]'.$$

Proof: See the Appendix. □

Drawing $\boldsymbol{\theta}$ from its marginal posterior increases the probability of acceptance substantially. Simulation studies performed on a small set of time points are encouraging; the parameter estimates obtained using these approximate marginal posteriors are not consequentially different from the estimates obtained without marginalizing over λ.

5.2. PSEUDO-IPM UPDATING AND APPROXIMATE POSTERIOR INFERENCE

As in Section 4, we consider x_j^* to be the center of the grid cell j. Then the "pseudo" IPM update is given by

$$\gamma_{t+1}\left(x_j^*\right) = \sum_l K_t\left(x_j^*|x_l^*; \mathbf{z}_t, \boldsymbol{\theta}, \gamma_{t,.}\right)\hat{\gamma}_t\left(x_l^*\right), \tag{5.4}$$

where $\hat{\gamma}_t(x)$ is an empirical estimate of the intensity corresponding to the point pattern observed at time point t evaluated at the grid centers x_l^*. Note that, under this updating scheme, for each t, we replace the t dimensional integral required to get $\gamma_t(x)$ in the original IPM update (2.3) by a one dimensional integral. Looking at (4.2), the term $[\{\gamma_t\}; \boldsymbol{\theta}]$ implied computing the γ_t deterministically and sequentially for a given $\boldsymbol{\theta}$, i.e., $\Pi_1^T[\gamma_t|\boldsymbol{\theta}, \gamma_{t-1}]$. Using (5.4), this now becomes $\Pi_1^T[\gamma_t|\boldsymbol{\theta}, \mathbf{x}_{t-1}]$ where $\mathbf{x}_{t-1}$ yields $\hat{\gamma}_{t-1}$. A graphical model may illuminate the difference. In Figure 1a, we show the graphical

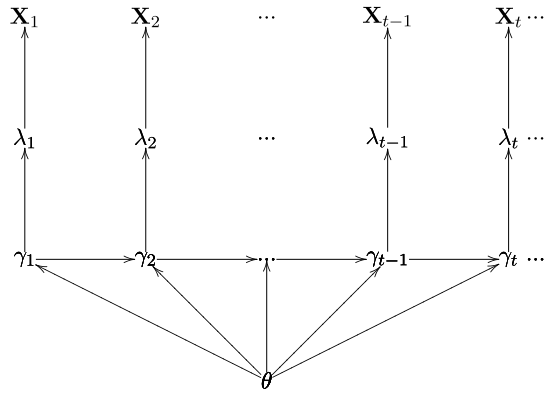

Figure 1a. Graphical model driving the updates in full IPM.

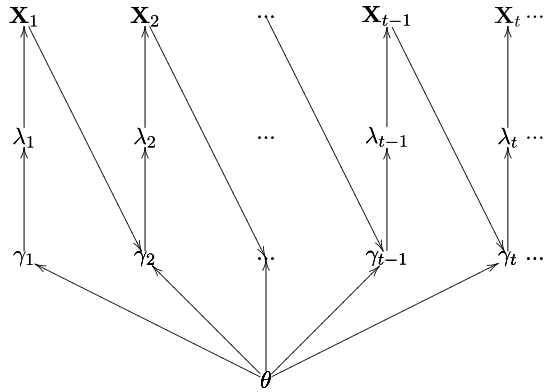

Figure 1b. Graphical model driving the updates in the "pseudo" IPM.

model associated with (4.2); in Figure 1b we show the revised graphical model using the "pseudo" IPM.

Thus, given $\boldsymbol{\theta}$, we can generate the sequence of $\gamma_t(x_j^*)$s with one-dimensional integrations using the "pseudo" IPM update (5.4). Additionally, the approximation simplifies the evaluation of the γ_t curves for any proposed value of $\boldsymbol{\theta}$ in the MCMC scheme. In the fitting, we gather the $\boldsymbol{\theta}$'s from each pair $\mathbf{x}_{t-1}, \mathbf{x}_t$, acknowledging that these samples of $\boldsymbol{\theta}$ do not come from the exact posterior distribution.

To remedy this problem, we propose to treat $\boldsymbol{\theta}$ as if they are time varying and evaluate the time-specific likelihood to draw the posterior samples of $\boldsymbol{\theta}$ for that time point. Let $\boldsymbol{\theta}_t = \{\boldsymbol{\theta}_{t1}, \ldots, \boldsymbol{\theta}_{tB}\}$ be the B samples of $\boldsymbol{\theta}$ obtained for time point t from the model $[\mathbf{x}_t|\boldsymbol{\lambda}_t][\boldsymbol{\lambda}_t|\{\gamma_t(x|\boldsymbol{\theta})\}][\boldsymbol{\theta}]$.[6] Corresponding to each $\boldsymbol{\theta}_t$, we have a sample of B $\{\gamma_t(x|\boldsymbol{\theta}_{tb})\}$s. The posterior mean of $\gamma_t(x|\boldsymbol{\theta})$ is given by $\hat{\gamma}_t(x) = \frac{1}{B}\sum_{b=1}^{B}\gamma_t(x|\boldsymbol{\theta}_{tb})$. We

[6]Here, for convenience we have switched the notation from the degenerate distribution $[\gamma_t; \boldsymbol{\theta}]$ to a degenerate distribution on the resulting function $\{\gamma_t(x|\boldsymbol{\theta})\}$.

plug $\hat{\gamma}_t(x)$ into the IPM update (2.3) to get $\gamma_{t+1}(.)$ and then proceed similarly to obtain the posterior samples, $\boldsymbol{\theta}_{t+1,b}$ at time $t+1$. Under the approximate marginalization described above, we essentially have samples from the following sequence of posteriors: $[\boldsymbol{\theta}|\mathbf{x}_1], [\boldsymbol{\theta}|\mathbf{x}_2, \hat{\gamma}_1(\mathbf{x}_1)], \ldots, [\boldsymbol{\theta}|\mathbf{x}_T, \hat{\gamma}_{T-1}(\mathbf{x}_{T-1})]$, where $\hat{\gamma}_t(x)$ is the posterior mean of $\gamma_t(x)$ obtained at time point t. But note that $\hat{\gamma}_{T-1}(\mathbf{x}_{T-1})$ is an implicit function of the entire sequence of the point pattern up to time $T-2$. So, at the last time point, the samples of $\boldsymbol{\theta}$ come from $[\boldsymbol{\theta}|\mathbf{x}_T, g(\mathbf{x}_{T-1}, \mathbf{x}_{T-2}, \ldots, \mathbf{x}_1)]$.

We seek samples of $\boldsymbol{\theta}$ from $[\boldsymbol{\theta}|\mathbf{x}_T, \ldots, \mathbf{x}_1]$. With samples of $\boldsymbol{\theta}$ for each time point, we can perform an importance sampling step to adjust for the approximate Laplace marginalization. Let $\boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{1B}, \boldsymbol{\theta}_{21}, \ldots, \boldsymbol{\theta}_{2B}, \ldots, \boldsymbol{\theta}_{T1}, \ldots, \boldsymbol{\theta}_{TB}$ be the samples of $\boldsymbol{\theta}$ obtained from the sequence of posteriors $[\boldsymbol{\theta}|\mathbf{x}_1], [\boldsymbol{\theta}|\mathbf{x}_2, \hat{\gamma}_1(\mathbf{x}_1)], \ldots, [\boldsymbol{\theta}|\mathbf{X}_T, g(\mathbf{x}_{T-1}, \ldots, \mathbf{x}_1)]$, respectively. Then, each $\boldsymbol{\theta}_{tb}$, $t = 1, \ldots, T$, $b = 1, \ldots, B$ is accepted with probability

$$w_{tb} = \frac{[\boldsymbol{\theta}_{tb}|\mathbf{x}_T, \ldots, \mathbf{x}_1]}{[\boldsymbol{\theta}_{tb}|\mathbf{x}_t]},$$

where $[\boldsymbol{\theta}_{tb}|\mathbf{x}_T, \ldots, \mathbf{x}_1] = \int_{\boldsymbol{\lambda}_T, \ldots, \boldsymbol{\lambda}_1} \prod_{t=1}^{T} [\mathbf{x}_t|\boldsymbol{\lambda}_t][\boldsymbol{\lambda}_t|\{\gamma_t(x|\boldsymbol{\theta}_{tb})\}][\boldsymbol{\theta}_{tb}]\, d\boldsymbol{\lambda}_1 \ldots d\boldsymbol{\lambda}_T$ and $[\boldsymbol{\theta}_{tb}|\mathbf{x}_t] = \int_{\boldsymbol{\lambda}_t} [\mathbf{x}_t|\boldsymbol{\lambda}_t][\boldsymbol{\lambda}_t|\{\gamma_t(x|\boldsymbol{\theta}_{tb})\}][\boldsymbol{\theta}_{tb}]\, d\boldsymbol{\lambda}_t$. Since both the marginalizations are done using Laplace approximation, the resulting ratio improves the order of the error (Tierney et al. 1989a, 1989b). Once we have the set $\{\boldsymbol{\theta}_{tb}, w_{tb}\}$, we can obtain the empirical distribution function, $\hat{F}_t(\boldsymbol{\theta})$, of $\boldsymbol{\theta}$ for every time point. These functions are then combined to give an estimate of the posterior distribution function of $\boldsymbol{\theta}$ as $\frac{1}{T}\sum_{t=1}^{T} \hat{F}_t(\boldsymbol{\theta})$.

5.3. PUTTING THE PIECES TOGETHER; POSTERIOR SUMMARIES

In order to assemble the ideas in the previous subsections into an overall model fitting procedure, we offer the following summary. Suppose a general version of K as in (4.3). That is, we have a model with q and Δ dependent upon size or with density dependence or both. Then, we confine ourselves to Figure 1b. In particular, we employ only the "pseudo" IPM, we revise the IPM component of the overall model as in Section 5.2, we use numerical integration as in (4.6), with a common $\boldsymbol{\theta}$ for all $t, t+1$ pairs.

With density independence, we have two paths. We can do exactly as with the density dependence case above. Alternatively, we can follow Figure 1a. That is, following Section 5.1, we use Fourier transformation with one dimensional inversion, followed by Laplace approximation to marginalize over λ_t. Under the marginalization, we introduce $\boldsymbol{\theta}_{tb}$ as in Section 5.2 and do the resampling and empirical distribution function computation as described there. Evidently, the posterior samples of $\boldsymbol{\theta}$ will be different according to the model we use. Assuming $[\boldsymbol{\theta}|\mathbf{x}_T, \ldots, \mathbf{x}_1] \approx [\boldsymbol{\theta}|\mathbf{x}_T, g(\mathbf{x}_{T-1}, \mathbf{x}_{T-2}, \ldots, \mathbf{x}_1)]$, we might choose to use the posterior samples of $\boldsymbol{\theta}$ obtained from the last time point to draw inference. If we want to use all the samples of $\boldsymbol{\theta}$ collected over all the time points, we resample them using w_{tb} as the weights and use the resampled $\boldsymbol{\theta}$s to draw inference. If we implement the "pseudo" IPM update, the posterior samples of $\boldsymbol{\theta}$ are obtained in a straightforward manner. Simulations in Section 6 under a known IPM suggest that the posterior inference for $\boldsymbol{\theta}$ obtained from these three strategies is very similar.

We also seek to draw inference on the intensity curves, in particular the γ_ts. To this end we proceed in the following manner. Let $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_B\}$ denote the posterior samples of $\boldsymbol{\theta}$. Then, for each $\boldsymbol{\theta}_b$, $b = 1, 2, \ldots, B$ we generate a sequence of $\tilde{\gamma}_t(x|\boldsymbol{\theta}_b)$, $t = 1, 2, \ldots, T$ using the recursion relation (5.2). Thus, the posterior mean of the curve γ_t is given by $E(\gamma_t(x)|\{\mathbf{x}_t, t = 1, 2, \ldots, T\}) \approx \frac{1}{B}\sum_{b=1}^{B} \tilde{\gamma}_t(x|\boldsymbol{\theta}_b)$. The upper and lower curves are obtained from the point-wise interval estimates. Estimates of the λ_t's may be of interest, in order to compare with the observed $\mathbf{x}_t$'s. These can be generated for each of the posterior samples of the $\boldsymbol{\theta}$'s and the $\boldsymbol{\gamma}_t$'s using the full conditional distribution which is proportional to the multivariate log-normal conditional distribution for $\boldsymbol{\lambda}_t|\boldsymbol{\gamma}_t$ times the contribution in the likelihood (using (4.6)) from $\mathbf{x}_t|\boldsymbol{\lambda}_t$.

6. EXAMPLES

6.1. SIMULATION STUDIES

In Section 6.1 we present two simulation examples to serve as a proof of concept. In Section 6.2 we consider the Duke Forest data.

6.1.1. A Density Independent IPM

Here, we simulate size distributions over time using a density independent IPM model. For convenience, we do not introduce covariates.

1. Set $L = 0.15$ and $U = 10$ and divide the interval into 100 grid cells of equal length. Let x_j^* denote the centers of the jth grid cell, $j = 1, \ldots, 100$.
2. Set $C_1^* = 50$, $\mu^* = 3$, $\sigma^{*2} = 1$ and obtain $\gamma_1(x_j^*)$ using the parametric form described below the expression in (5.1).
3. Obtain $\boldsymbol{\lambda}_1$ using the measurement error model $\lambda_1(x_j^*) = \gamma_1(x_j^*) \times \epsilon(x_j^*)$, where $(\epsilon(x_1^*), \ldots, \epsilon(x_{100}^*)) \sim$ Lognormal$(0, \Sigma)$. As described above, Σ has the exponential covariance structure with $\sigma_\epsilon^2 = 0.01$ and $\phi = 0.9$.
4. Define $\tilde{\lambda}_1 = \max\{\lambda_1(x_1^*), \ldots, \lambda_1(x_{100}^*)\} + 1$ and generate a point pattern from a homogeneous Poisson process with intensity $\tilde{\lambda}_1$.
5. Using the ratio $\frac{\lambda_1(x^*)}{\tilde{\lambda}_1}$, we thin the above point pattern to arrive at our realized point pattern for time point 1.
6. Fix $\mu_t = 1, \sigma_t^2 = 0.8, q_t = 0.9, \eta_t = 1, \Delta_t = 0.1, \forall t$. Plug in these parameters in the IPM update (2.3) to get $\gamma_2(x_j^*)$ and repeat Steps 3, 4 and 5 to obtain a point pattern realization corresponding to time point 2.
7. Using this scheme, we generate marginal point patterns for five consecutive years.

In fitting the IPM, we assume all the parameters in $K(.)$ are global and remove the hierarchical level that relates the parameters of $K(.)$ to the covariates using the regression hyper-parameters. As described earlier, we set μ^* and σ^{*2} to be the mode$(\mathbf{x}_1)$

Table 1. Comparing the estimates of the global parameters obtained using two computational strategies (see Section 6.1.1).

Parameters	True value	Posterior summary of $\boldsymbol{\theta}_T$	Posterior summary of $\boldsymbol{\theta}_I$	Posterior summary of $\boldsymbol{\theta}_P$
q	0.9	0.83 (0.76, 0.95)	0.83 (0.76, 0.92)	0.78 (0.69,0.94)
μ	1	1.03 (0.73, 1.44)	1.05 (0.74, 1.41)	0.89 (0.73, 1.56)
σ^2	0.8	0.72 (0.61, 1.01)	0.66 (0.47, 0.95)	0.71 (0.53, 1.21)
η	1	0.93 (0.75, 1.22)	0.94 (0.78, 1.23)	0.87 (0.68, 1.29)
Δ	0.1	0.08 (0.005, 0.22)	0.08 (0.003,0.17)	0.09 (0.005,0.27)
σ_ϵ^2	0.01	0.017 (0.005, 0.028)	0.005 (0.001, 0.026)	0.016 (0.001, 0.034)
Computation times	–	3 hrs 12 mins	3 hrs 25 mins	1 hr 05 mins

Table 2. Comparing the MSEs obtained using two computation strategies (see Section 6.1.1).

Time	MSE using $\boldsymbol{\theta}_I$	MSE using $\boldsymbol{\theta}_P$
2	1.83	2.74
3	1.81	2.94
4	1.04	2.44
5	1.23	2.04

and var($\mathbf{x}_1$), respectively. The priors on the parameters of the redistribution kernel are as follows: $\mu_t = \mu \sim \text{Normal}(0, 10^2)$, $\sigma_t^2 = \sigma^2 \sim \text{IG}(10, 2)$, $q_t = q \sim \text{Uniform}(0, 1)$, $\Delta_t = \Delta \sim \text{Uniform}(0, 1)$, $\eta_t = \eta \sim \text{lognormal}(0, 10^2)$. The prior for σ_ϵ^2 is developed in the following way: For each year we estimate the empirical intensity, $\tilde{\boldsymbol{\lambda}}_t$. An initial estimate of $\boldsymbol{\theta}$ is obtain by setting $\tilde{\boldsymbol{\theta}} = \text{argmin}_{\boldsymbol{\theta}} \sum_t (\tilde{\boldsymbol{\lambda}}_t - \boldsymbol{\gamma}_t(x|\boldsymbol{\theta}))^2$. Using $\tilde{\boldsymbol{\theta}}$, we generate the whole sequence of $\{\boldsymbol{\gamma}_t(x|\tilde{\boldsymbol{\theta}})\}$. Then, define $\tilde{\mathbf{V}}_t = \text{var}(\log(\tilde{\boldsymbol{\lambda}}_t) - \log(\boldsymbol{\gamma}_t(x|\tilde{\boldsymbol{\theta}})))$. The prior of σ_ϵ^2 is taken to be inverse gamma with mean $\sum_t \tilde{\mathbf{V}}_t/t$ and variance var($\tilde{\mathbf{V}}_1, \ldots, \tilde{\mathbf{V}}_t$).

Posterior inference is drawn using (i) only the samples generated at the last time point, labeled $\boldsymbol{\theta}_T$, (ii) using the importance sampling scheme in Section 5.2 to retain samples for each of the t's, labeled $\boldsymbol{\theta}_I$ and (iii) the samples generated using the "pseudo" IPM scheme with density dependence, labeled $\boldsymbol{\theta}_P$. Table 1 shows the comparison of estimates of $\boldsymbol{\theta}$ obtained from these three different schemes and also compares the computation time required for each scheme to generate 50,000 samples under MCMC, with the first 10,000 discarded as burn-in. The resampled $\boldsymbol{\theta}$s yield almost identical results to those obtained by using the samples of $\boldsymbol{\theta}_T$. Moreover, the density-dependent IPM gives reasonable estimates compared to those obtained from the other two schemes even though this is done under misspecification. In Figure 2, we plot the posterior mean curve of γ_t obtained from the posterior samples $\boldsymbol{\theta}_I$ and $\boldsymbol{\theta}_P$ against the true γs used to simulate the data for three of the five time points. The solid line shows the true simulated γ_t, the dotted line shows the γ_ts estimated using $\boldsymbol{\theta}_I$s and the dashed line shows the γ_t obtained using $\boldsymbol{\theta}_P$. Since we know the true intensity, we can quantify how the estimated intensities differ from the true one by using the Mean Integrated Square Error criterion (discretized using the 100 x_j^*'s). Table 2

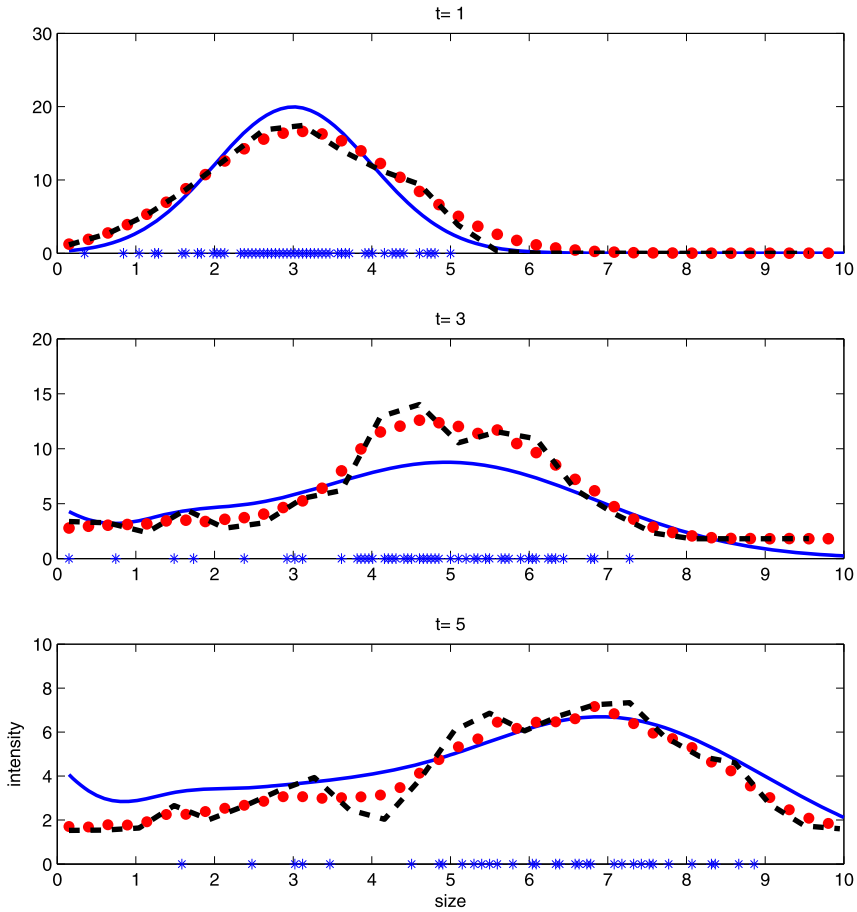

Figure 2. Comparing the estimated γ_ts using the importance sampling scheme and "pseudo" IPM scheme (Section 6.1.1). The true simulated γ_ts are shown with solid line. The γ_ts estimated using importance sampling scheme are shown with dotted line, the dashed line shows γ_ts obtained using the "pseudo" IPM scheme. The rug stars (*) show the simulated data points.

shows the MISE's obtained using $\boldsymbol{\theta}_I$ and $\boldsymbol{\theta}_P$, starting from the second time point. We see that, not surprisingly, the $\boldsymbol{\theta}_I$ do better since the true model is not density dependent.

6.1.2. An IPM with Size Dependence

Here, we make the redistribution kernel, K, more flexible by including size information in the survival probabilities $q_t(x)$ and making the growth density asymmetric. We describe the simulation procedure below:

Once again we assume the lower bound (L) to be 0.15 and the upper bound (U) to be 10. The model for survival probability is given by

$$q(x;\boldsymbol{\beta}_q) = \frac{\exp(\beta_{0q} + \beta_{1q}x)}{1 + \exp(\beta_{0q} + \beta_{1q}x)}.$$

We specify $\beta_{0q} = 2.26$ and $\beta_{1q} = 0.23$. These values of β_{0q} and β_{1q} correspond to a minimum survival probability of 0.91 corresponding to the smallest size class and a maximum survival probability of 0.99 corresponding to the largest size class. The growth kernel, $f(y|x)$ is given by $\frac{1}{y\sigma_f\sqrt{2\pi}} \exp(-\frac{(\log(y)-\log(x))^2}{2\sigma_f^2})$ so the expected value of $y|x$ is given by $x \exp(\sigma_f^2/2)$. In order to *distribute* new y's roughly around a current x, we assume a small value for $\sigma_f^2 = 0.5$. We do not assume any size dependence in the fecundity part. For simplicity, we keep the expected recruitment per individual (Δ) constant at 0.15. The η in the recruitment size density ($\eta \exp(-\eta(y - L))$) is kept fixed at 1.5.

We assume Uniform priors on β_{0q} and β_{1q} with informative support. The lower and upper bounds for the priors of these two parameters are fixed such that the lowest survival probability corresponding to the lowest size class does not go below 0.80 and the highest survival probability corresponding to the highest size class does not exceed 0.9999. Satisfying these conditions, we assume a Uniform(1.38, 2.5) prior on the former and a Uniform(0.05, 0.5) prior on the latter. For σ_f^2, we assume a Gamma prior with mean 1 and variance 2. As discussed in Section 2.3, we anticipate $q(x) + \Delta \in (0.9, 1.1)$, so we assume a Uniform(0.1, 0.3) prior on Δ. For η, we expect that the new recruits will occupy the first few size classes and hence specify a Uniform(0.5, 5) prior on it. Table 3 shows the posterior mean and 95 % CI (in parentheses) of the parameters obtained from fitting the pseudo IPM. Figure 3 shows the true simulated $\gamma_t(x)$ and the estimated mean log intensity along with the simulated point pattern for three different time points.

6.1.3. A Density Dependent IPM

Next, we turn to a density dependent example using (4.4) and (4.5). To attain stability of the population sizes in the density dependent IPM, we assume that we have an upper bound for the survival function $q(.)$ and the recruitment function $\Delta(.)$ denoted by q_u and Δ_u, respectively. These are achieved when $\gamma_{t,.}$ is 0 and yield the values of Q_0 and δ_0 using the equations:

$$\frac{Q_0}{1 + Q_0} = q_u, \tag{6.1}$$

and

$$\frac{\delta_0}{1 + \delta_0} = \Delta_u. \tag{6.2}$$

In Figure 4 we show the behavior of population sizes for different values of q_u, Δ_u, Q_1 and δ_1. The top panel shows a scenario where population increases over time until it reaches a stable limit. Such a situation could be arrived at by fixing $q_u = 0.97$, $\Delta_u = 0.2$, $Q_1 = \delta_1 = 0.009$. A population extinction scenario (middle panel) can be obtained when we fix $q_u = 0.96$, $\Delta_u = 0.02$, $Q_1 = 0.009$ and $\delta_1 = 0.91$. The bottom panel shows an oscillating population size which eventually reaches a stable limit. Such an oscillating scenario can be obtained when $q_u = \Delta_u = 0.97$, $Q_1 = 0.1$ and $\delta_1 = 0.15$. Because $q(.)$ and $\Delta(.)$ are strictly decreasing, we can never have explosive population size.

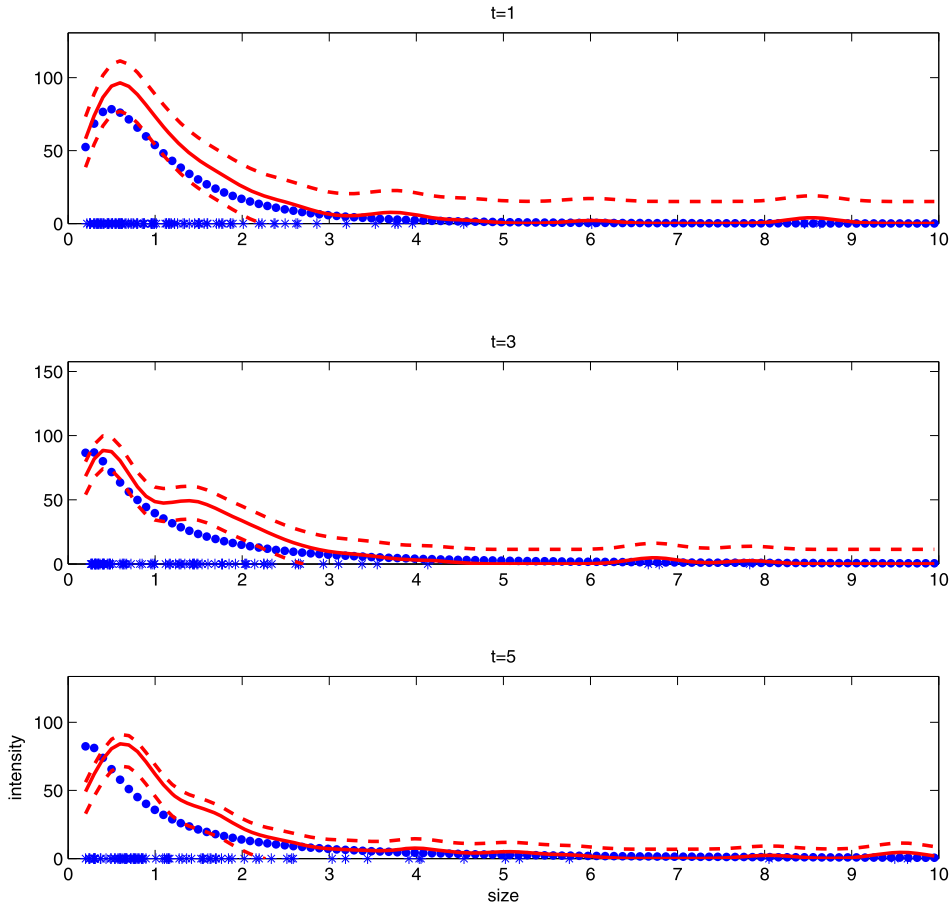

Figure 3. Plot of true γ_ts (dotted line) and estimated γ_ts (solid line) using the size-dependent IPM (Section 6.1.2), along with the pointwise 95 % CI (dashed line) estimates for three time points. The rug stars (*) show the simulated point pattern of sizes obtained at the corresponding time points.

Table 3. Model retrieval under size dependent IPM (see Section 6.1.2).

Parameters	True value	Posterior summary under stable population
β_{0q}	2.26	2.12 (1.76, 2.39)
β_{1q}	0.23	0.35 (0.08, 0.42)
σ_f^2	0.50	0.70 (0.23, 1.85)
η	1.50	1.94 (1.004, 2.23)
Δ	0.15	0.12 (0.03,0.24)
σ_ϵ^2	0.01	0.004 (0.0001, 0.035)

To illustrate the performance of the density dependent IPM, we use the population scenario described in the top panel of Figure 4. We randomly generate an initial population of size 10 in the interval [0.15, 3]. We set $Q_0 = 31.69$, $Q_1 = 0.009$ and $\delta_0 =$

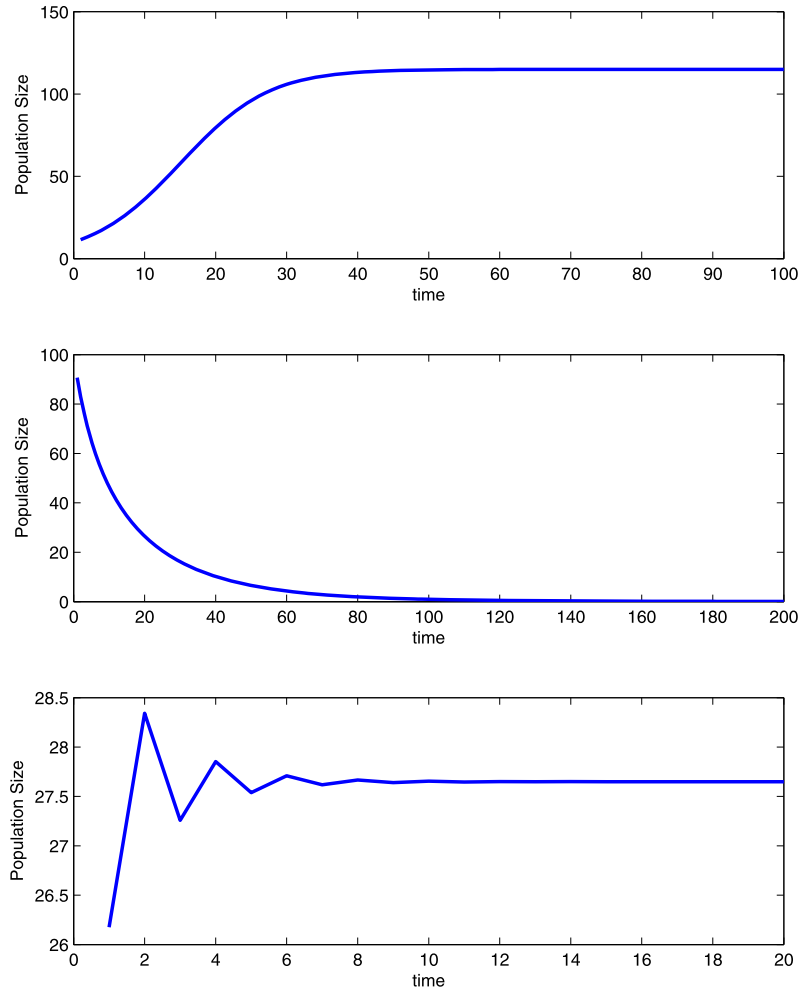

Figure 4. Three population size scenarios using density dependent IPM: population size increasing monotonically over time until it reaches a stable size (top panel); a population extinction scenario (middle panel); population size oscillating over time until it reaches a stable size (bottom panel).

0.25, $\delta_1 = 0.009$. Then $q(\gamma_{t,\cdot})$ and $\Delta(\gamma_{t,\cdot})$ are given by Equations (4.4) and (4.5). We fix $\mu_t = \mu = 0.1, \sigma_t^2 = \sigma^2 = 0.2, \eta_t = \eta = 1$ and, using the density dependent survival probability and expected influx, we generate γ_{t+1} from (2.3). Then, taking $\sigma_\epsilon^2 = 0.01$, we simulate the point pattern of sizes using the procedure described above. Figure 5 shows the simulated $\gamma_{t,\cdot}$ and $\lambda_{t,\cdot} = \int_L^U \lambda_t(x)\,dx$ over time. We see that the population grows initially and then stabilizes starting roughly at $t = 40$.

Although we have simulated marginal point patterns for 100 time points, such rich data may not be available in practice. So we train the density-dependent IPM on an initial portion of this simulated data and forecast $\gamma_t(x)$ for the rest of the time points. We then compare the predicted intensity with the true one to assess the adequacy of such forecasts. While fitting the model, we use the "pseudo" IPM procedure described in (5.4). We assume

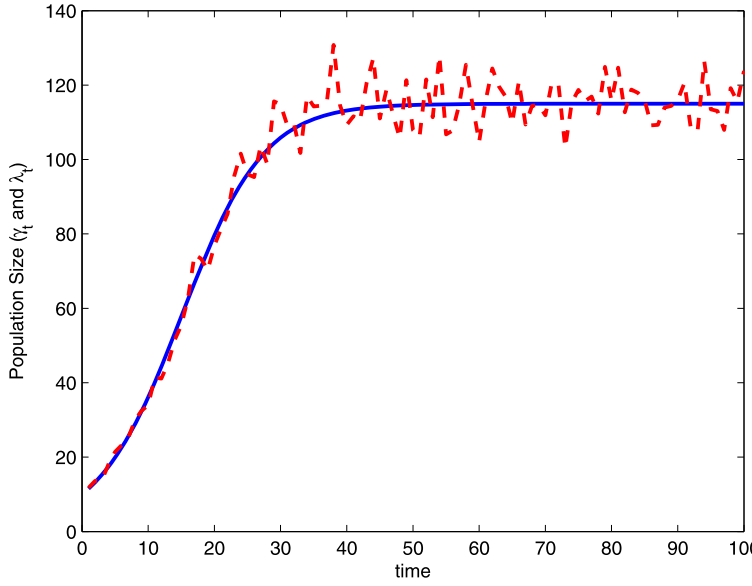

Figure 5. The simulated $\gamma_{t,.}$ (solid line) and $\lambda_{t,.}$ (dashed line) for 100 time points (Section 6.1.3).

that q_u and Δ_u are known and then solve (6.1) and (6.2) to get Q_0 and δ_0, respectively. Since we have no additional information regarding the values of Q_1 and δ_1, we estimate them as part of the model fitting. However, given the form of $q(\gamma_{t,.})$ and $\Delta_{\gamma_{t,.}}$ and that they enter additively in $\int K(y;x)\,dy$, Q_1 and δ_1 are not identifiable unless we have strong prior information. However, the sum is. The prior choices are same as described in the previous section and we employ a Uniform(0, 0.2) prior on Q_1 and δ_1.

As the first exercise, we train the model on the first 20 time points and forecast $\gamma_t(x)$ on the remaining 80 time points. Figure 6a shows the plot of the true simulated γ versus the estimated one along with the lower and upper curves, obtained from the pointwise 95 % interval estimates for time points 1, 20, 35, 100. Clearly, the uncertainty increases as the forecast horizon increases. Next, we train the model on the first 30 time-points and forecast $\gamma_t(x)$ on the remaining time-points. Figure 6b shows the plot of the true simulated γ versus the estimated one along with the lower and upper curves, obtained from the pointwise 95 % interval estimates, again for time points 1, 20, 35, 100. Once again, the uncertainty increases as the forecast horizon increases, but with the additional data, the estimated $\gamma_t(x)$ are closer to the true values and the uncertainty bands are tighter than what we observe in Figure 6a. Figure 7a shows the plot of $\gamma_{t,.}$ estimated using the first 20 time points versus the true $\gamma_{t,.}$. In Figure 7b, we plot $\gamma_{t,.}$ estimated using the first 30 time points against the true $\gamma_{t,.}$. In both figures we show the 45° line (in solid) to visually ascertain how well each model can capture the dynamics of the population size. Since $\gamma_{t,.}$ increases monotonically in this simulation, one can infer that the forecasts worsen for both models as we predict deeper into future. However, one can also see that $\gamma_{t,.}$ estimates obtained using 30 time

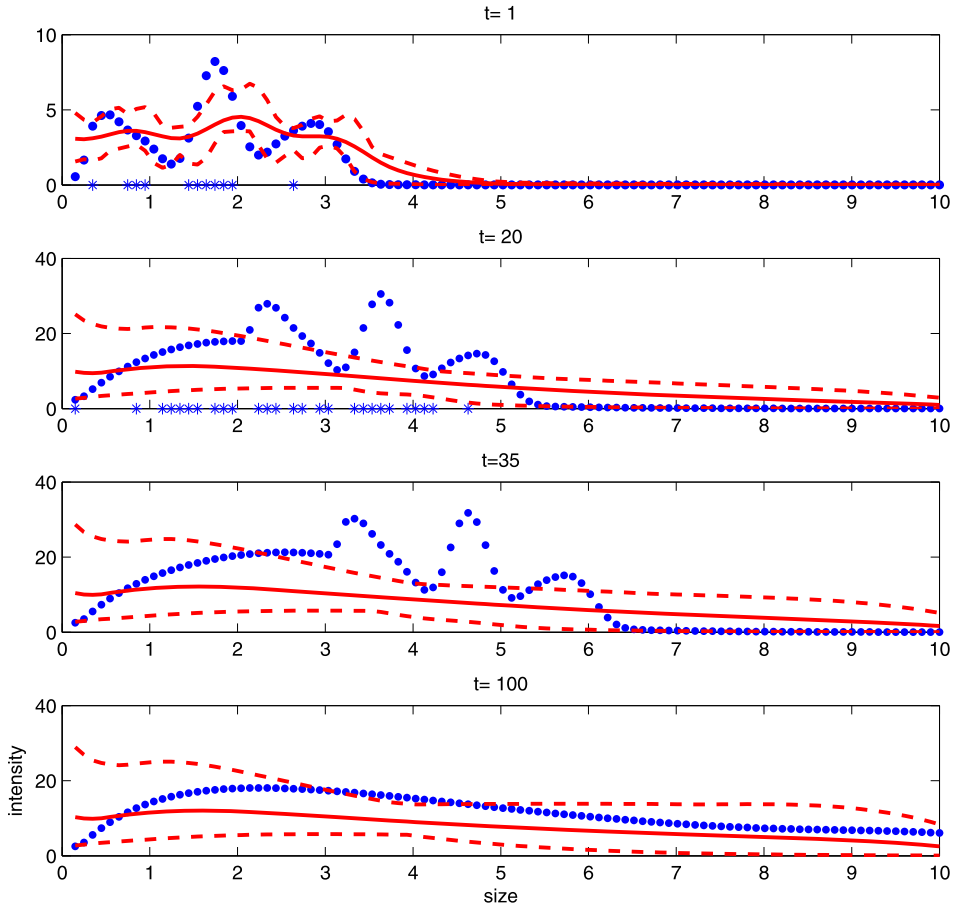

Figure 6a. Plot of true simulated γ_ts (dotted line) and estimated γ_ts (solid line), for the density-dependent IPM, along with the pointwise 95 % CI (dashed line) estimates using data from the first 20 time-points.

points show considerable improvement as compared to the estimates obtained when we use only 20 time points. Table 4 shows the posterior summary of the parameters estimated using the simulated data for 30 time points.

6.2. TWO ANALYSES OF DUKE FOREST DATA

Here, we consider two analyses using Duke Forest data. In Section 6.2.1, we implement a full IPM fitting using our modeling approach. In Section 6.2.2 we look at a comparison between projections using our approach and those from fitting with individual-level data.

6.2.1. Our Modeling Approach Applied to Duke Forest Data

Measurements of stem diameter of the tree species *Liriodendron tulipifera* (Litu), observed in the Blackwood site of the Duke Forest, were obtained from 2000 to 2010. The number of trees varied from 305 in 2004 to 340 in 2010. The initial population size in

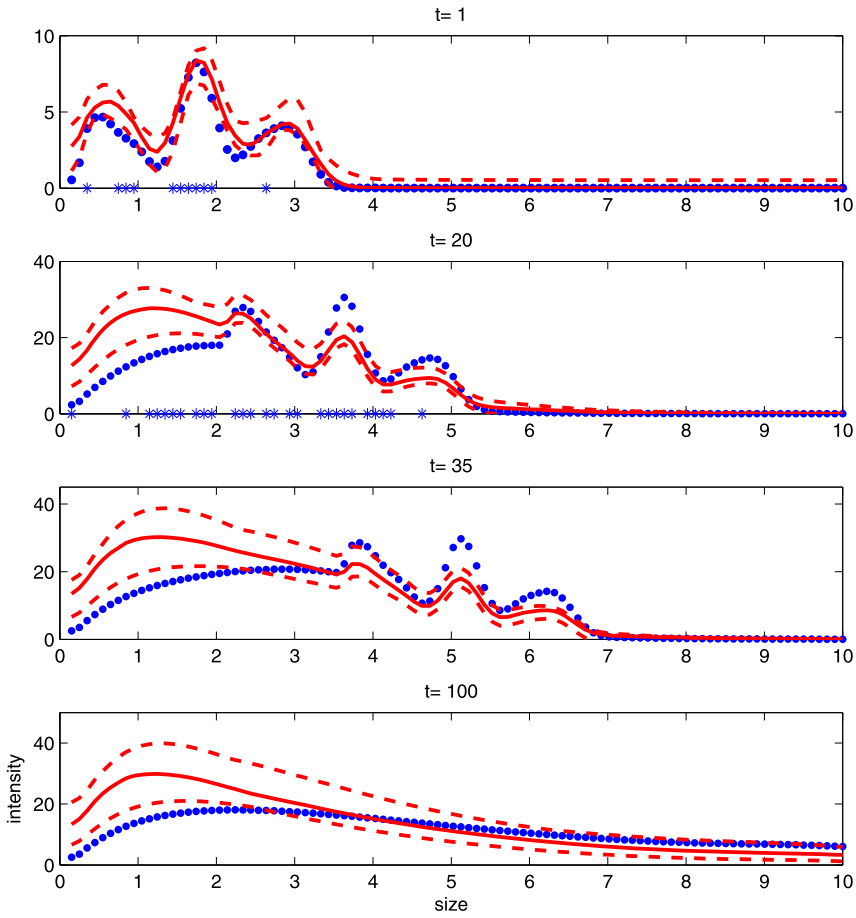

Figure 6b. Plot of true simulated γ_ts (dotted line) and estimated γ_ts (solid line), for the density-dependent IPM, along with the pointwise 95 % CI (dashed line) estimates using data from the first 30 time-points.

Table 4. Model retrieval under density dependent IPM (see Section 6.1.3).

Parameters	True value	Posterior summary under stable population
μ	0.1	0.16 (0.06, 0.23)
σ^2	0.2	0.15 (0.09, 0.35)
η	1	1.16 (0.71, 1.28)
Q_1	0.009	0.007 (0.0005, 0.017)
δ_1	0.009	0.012 (0.007,0.028)
σ^2_ϵ	0.01	0.008 (0.005, 0.026)

year 2000 was 317. Over the course of the data collection, diameters ranged from 0.153 to 82.375 cm. As covariates, we used winter temperature (degrees centigrade) and summer drought, as captured through the Palmer Drought Severity Index (PDSI), centered

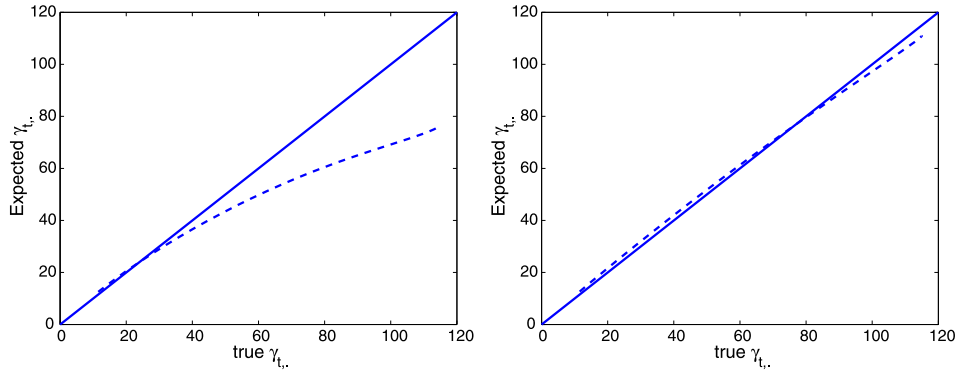

Figure 7. (a) Plot of true simulated $\gamma_{t,.}$ and estimated $\gamma_{t,.}$ (dashed line), for the density-dependent IPM, using data from the first 20 time-points along with 45° line (solid) and (b) plot of true simulated $\gamma_{t,.}$ and estimated $\gamma_{t,.}$ (dashed line), using the density-dependent IPM, using data from first 30 time-points along with 45° line (solid).

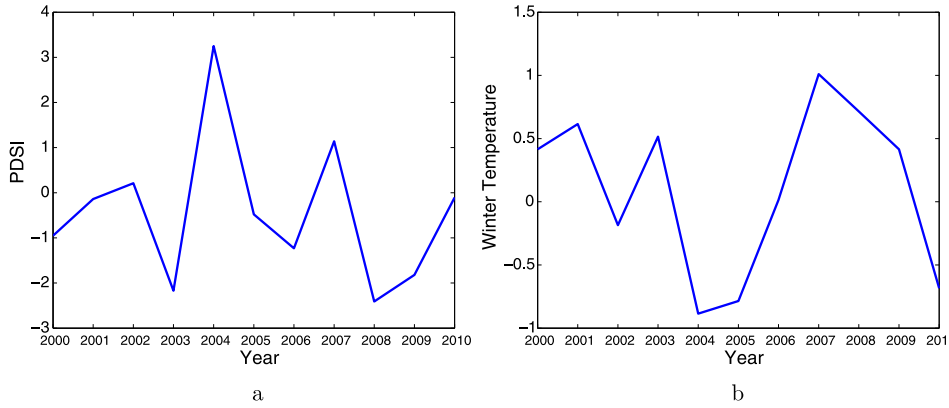

Figure 8. (a) Plot of the PDSI versus year and (b) plot of the winter temperature versus year.

and scaled about their sample mean and sample standard deviation, respectively (Clark et al. 2011a, 2011b). In Figures 8a and 8b, respectively, we plot the raw PDSI and winter temperature observed during the study period, revealing considerable annual variation.

Because large changes in population size are not expected over a single decade (not revealed by the data), we fitted the density-independent IPM using Fourier transformation with one dimensional inversion, followed by Laplace approximation to marginalize over λ_t. We use the regression models for the vital rates in K described above, including a logistic regression for q and Δ with no population size ($\gamma_{t,.}$) term. The final estimates are obtained using the resampled $\boldsymbol{\theta}^*_{tj}$ following the procedure outlined in Section 5.2. We use the priors specified in Section 5.1 and set the lower limit L and upper limit U at 0.10 and 100 cm, respectively. We hold out the data from year 2010 and train the model on the earlier time points. Then, we use the covariates for the validation year and predict $\gamma_t(x)$

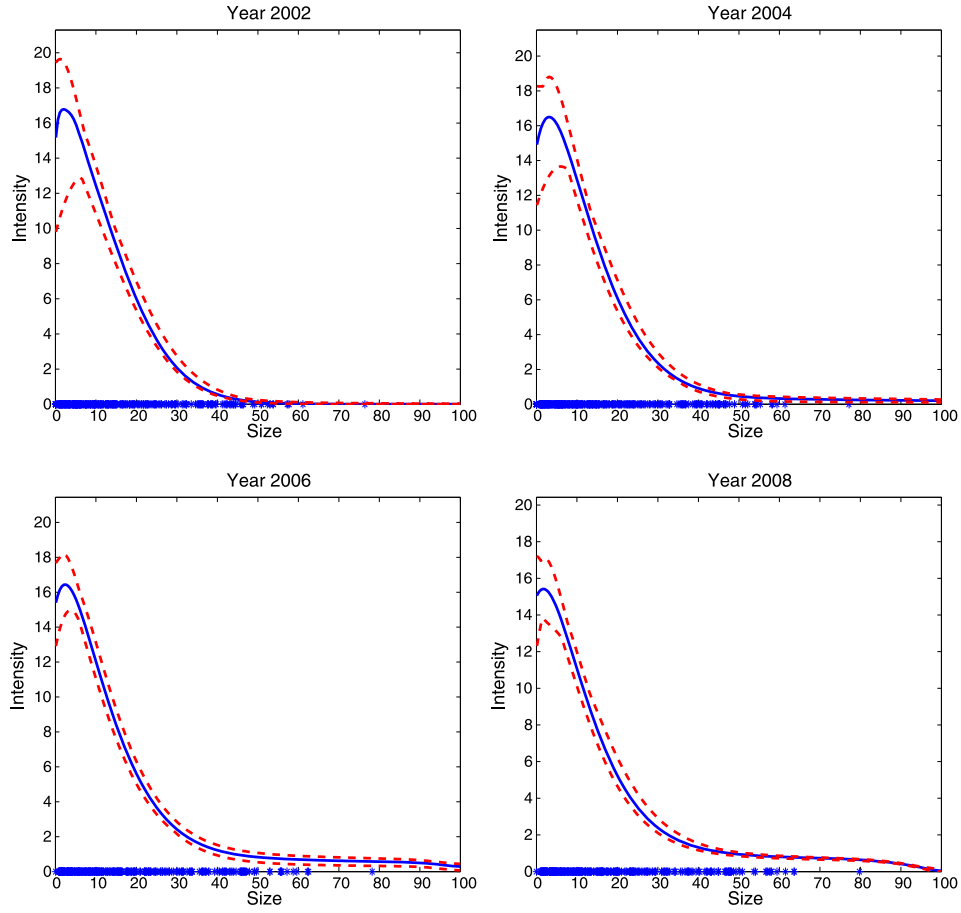

Figure 9a. Plot of $\gamma_t(x)$ (solid line) along with the pointwise 95 % CI (dashed line). The rug stars (*) show the observed size patterns.

for that year. Table 5 provides a posterior inference summary for the model parameters. Figure 9a shows the posterior mean curve of γ_t along with the 95 % pointwise interval estimates for the data in the training sample for even years beginning in 2002. Figure 9b shows the predicted $\gamma_t(x)$ for 2010 along with the uncertainty associated with it.

Figure 10a shows the posterior summary of the survival probability, q, and Figure 10b shows the expected influx, Δ, for each year. Finally, Figure 11a shows the posterior mean survival probability surface $(\frac{1}{B}\sum_{l=1}^{B}\frac{\exp(\mathbf{z}\boldsymbol{\beta}^*_{q,l})}{1+\exp(\mathbf{z}\boldsymbol{\beta}^*_{q,l})}$, where $\boldsymbol{\beta}^*_{q,l}$ are the posterior samples of $\boldsymbol{\beta}_q)$ over $\mathbf{z}$. Figure 11b shows the posterior mean influx per individual surface $(\frac{1}{B}\sum_{i=l}^{B}\frac{\exp(\mathbf{z}\boldsymbol{\beta}^*_{\Delta,l})}{1+\exp(\mathbf{z}\boldsymbol{\beta}^*_{\Delta,l})}$, where $\boldsymbol{\beta}^*_{\Delta,l}$ are the posterior samples of $\boldsymbol{\beta}_\Delta)$ over $\mathbf{z}$. To demonstrate the short-term predictive performance of our IPM model, we consider year-to-year changes in population sizes, plotting the observed and expected (under our IPM) population sizes in Figure 12. The curves are in close agreement.

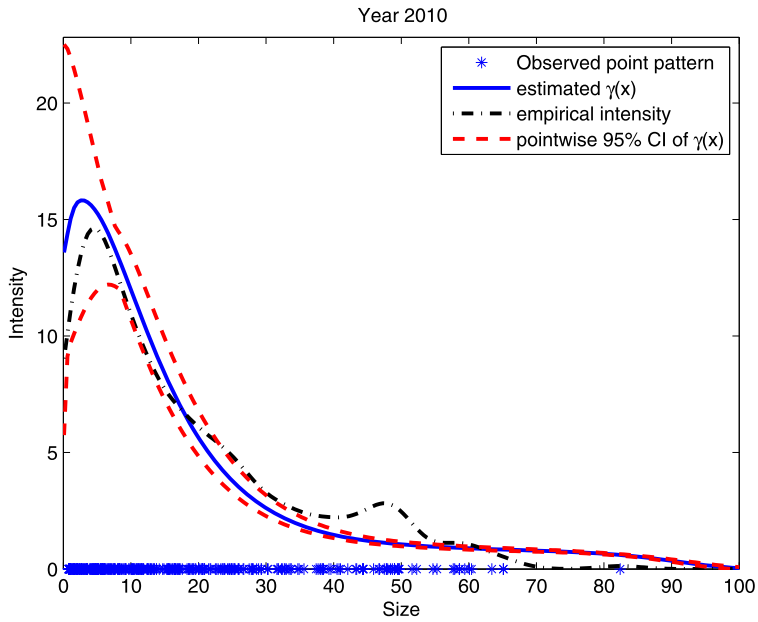

Figure 9b. Plot of the predicted $\gamma_t(x)$ (solid line) along with the pointwise 95 % CI for year 2010 (dashed line). The rug stars (*) show the observed size pattern. The empirical intensity corresponding the observed point pattern is shown with (-.).

Table 5. Posterior summary of the regression parameters governing the size distribution of Litu using density-independent IPM (see Section 6.2.1).

Parameters	Posterior summary
β_q Intercept	2.24 (1.05, 2.48)
β_q PDSI	−0.15 (−0.45, 0.17)
β_q Wtemp	0.14 (−0.30, 0.37)
β_μ Intercept	−2.43 (−3.15, −1.84)
β_μ PDSI	0.17 (−0.31, 0.96)
β_μ Wtemp	0.19 (−0.43, 0.63)
β_{σ^2} Intercept	2.06 (1.36, 2.72)
β_{σ^2} PDSI	−0.15 (−1.19, 0.56)
β_{σ^2} Wtemp	−0.23 (−0.86 1.04)
β_Δ Intercept	−3.00 (−4.01, −2.46)
β_Δ PDSI	−0.44 (−1.41, −0.01)
β_Δ Wtemp	−0.68 (−1.91, −0.19)
β_η Intercept	−2.56 (−3.01, −2.98)
β_η PDSI	0.78 (0.51, 1.00)
β_η Wtemp	0.79 (0.42, 0.96)
σ^2_ϵ	0.04 (0.002, 0.1)

6.2.2. Comparing Projections

Comparing projections under an IPM fitted with individual data vs. one fitted using our approach is difficult. If we only have point patterns over time, we cannot fit the individual

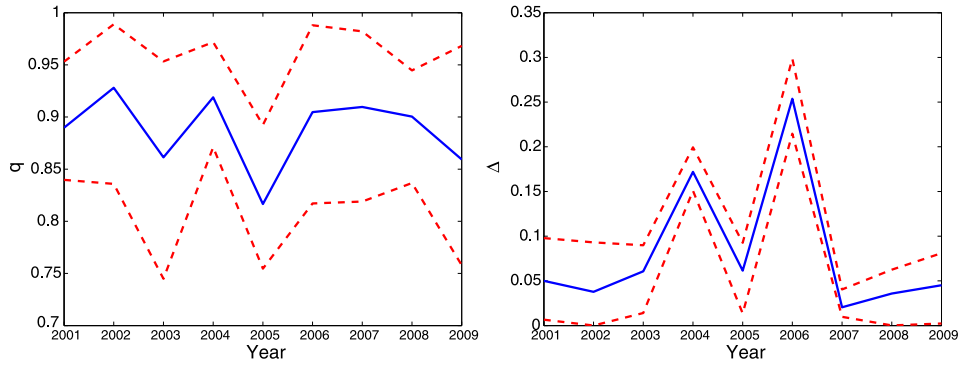

Figure 10. (a) Posterior mean of the survival probability, q_t (solid line) along with the pointwise 95 % CI (dashed line) and (b) posterior mean of the influx per individual, Δ_t (solid line) along with the pointwise 95 % CI (dashed line)

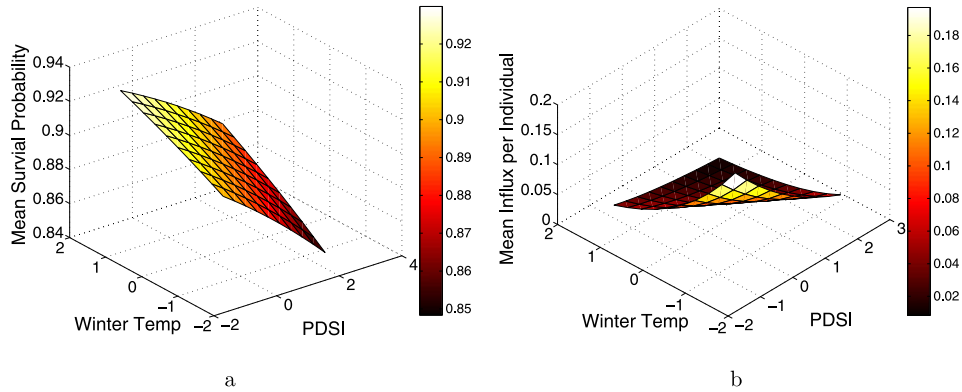

Figure 11. (a) Posterior mean survival probability surface over the covariate space and (b) posterior mean influx per individual surface over the covariate space.

level model; with individual level data, it is unclear what sort of dynamic point patterns will arise and whether our intensity-based modeling assumptions apply. Nonetheless, we supply an example which reveals the advantage for short term prediction when fitting K using our modeling approach compared with fitting using individual level data and then projecting. We use the foregoing Duke Forest data to make this comparison. However, we remove the time dependence in K by removing both density dependence and climate dependence, using only size dependence. Moreover, we only compare observed annual population size with the population size estimated under the two models.

From Rees and Ellner (2009), the long term growth rate is given by $E(\log(N(t+1)/N(t)))$ (assuming this expectation exists). Under appropriate assumptions, we can invoke the law of large numbers and the long term growth rate can be written as

$$\lim_{t\to\infty} \frac{1}{t} \sum_{j=0}^{t-1} \log\left(\frac{N(j+1)}{N(j)}\right) = \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{N(t)}{N(0)}\right). \tag{6.3}$$

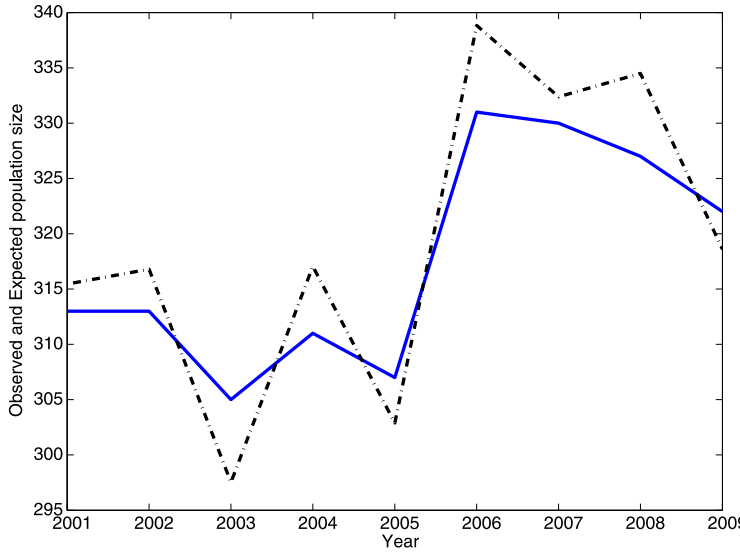

Figure 12. Plot of observed population sizes (solid line) and the expected population sizes (dash-dotted line) obtained under our IPM from year 2001 through 2009.

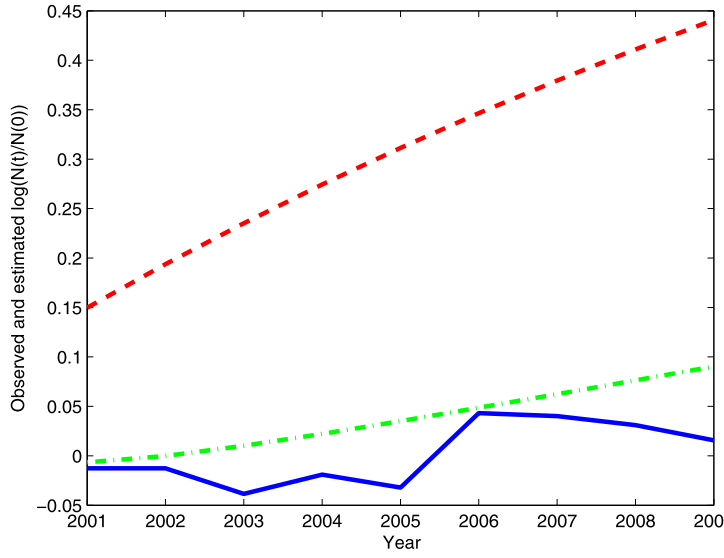

Figure 13. Plot of $\log(\frac{N(t)}{N(0)})$ with respect to time (solid line). Overlaid are the plots of $E(\log(\frac{\gamma_{t,.}}{\gamma_{0,.}})|\text{Data})$ (dash-dotted line) and the projected population estimate obtained from individual based IPM (dashed line). Both the IPMs use stationary, size-dependent redistribution kernel.

The empirical estimate of (6.3) is $\frac{1}{T}\log(\frac{N(T)}{N(0)})$, where T is the last time point.[7] For the Litu data, this value is 0.0016. For our IPM model, we can compare this with the posterior

[7]As T grows large, the effect of normalizing by $N(0)$ diminishes so, often the growth rate is written without normalizing by $N(0)$ as in Ellner and Rees (2007).

mean of $\frac{1}{T}\log(\frac{\gamma_{T,\cdot}}{\gamma_{0,\cdot}})$, which is 0.009 (again $\gamma_{0,\cdot}$ is an empirical intensity). The 95 % credible interval for this quantity is $(-0.015, 0.021)$ which contains the empirical estimate. In fact, Figure 13 plots the empirical estimates $\log(\frac{N(t)}{N(0)})$ along with our corresponding posterior mean for $t = 1, 2, \ldots, 9$. Given the foregoing simplified model assumptions, the curves are in reasonable agreement.

For the individual data based IPM, working with trees, often only some of the vital rates can be estimated, in which case ad hoc assumptions are used to fill out the kernel. We applied an approach similar to Zuidema et al. (2010) for tree population dynamics, where, as in our situation, they could not actually observe fecundity and thus used aggregate values for transitions from trees to seedlings. We constructed transition kernels for growth of seedlings and trees using similar assumptions, i.e., fitted GLMs for adults and aggregate transitions for recruitment. We evaluated eigenvalues from discretized kernels as is standard practice with IPMs. Because we do not have initial distributions for all seedling classes we projected the population forward from the initial tree size distribution (individuals larger than 2 meters tall). Without input of new individuals across the L boundary, densities rapidly decay. We used the same assumptions about aggregate behavior in model fitting to evaluate flux across the L boundary, based on growth rates of seedlings. Then $\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t) + \mathbf{v}$, where $\mathbf{v}$ is a vector of zeros except for the first class, which contains the input from seedling growth into the first diameter size class. The resulting numbers enable estimates relative to the observed $N(0)$ to compare with the empirical growth rates. They are overlaid in Figures 13. The estimated growth rate obtained from this individual level model is 0.044. With regard to short term prediction, we see substantial discrepancy between the inference from the individual level model and our model as well as the empirical estimates. Unfortunately, when fitting the individual level IPM, there is no uncertainty that we can attach to the dashed line arising from that model so we cannot make formal inferential comparison with the empirical plot.

7. SUMMARY AND FUTURE WORK

We have presented a new view for specifying, fitting, and analyzing population dynamics using IPM models. We can accommodate both density independent and dependent cases and incorporate size dependence. We work within a three-stage Bayesian hierarchical modeling framework, resulting in very demanding model fitting. We have offered several approximate fitting strategies which, based upon simulation study, seem to perform well. We also presented the results of fitting an IPM for size distribution of a tree species using data from Duke Forest.

Future work will see us investigating additional species to compare IPM's. We will also explore the possibility of building a joint IPM specification to allow for dependence between species. Such dependence could obviously affect both population size and size distribution. We will also turn to the U.S. Forest Service's Forest Inventory Analysis dataset. This dataset spans more than 42,000 plot measurements over the eastern U.S. but samples only 20 % each year. Two major related questions emerge: how can we fit an IPM when we never observe consecutive years of data for any plot and how can we scale an IPM to

accommodate such a large region? We will offer approaches to address these questions in future work.

[Published Online December 2012.]

APPENDIX

Approximate Inverse Fourier Transform The inverse Fourier transform (IFT) is a one-dimensional integral given by

$$\gamma_t(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-iux}\tilde{\gamma}_t(u)\,du. \tag{A.1}$$

We approximate this IFT as follows:

1. At the outset we assume that $\gamma_t(x)$ is effectively 0 outside the interval $[0, A]$ (where $[L, U] \subset [0, A]$).
2. Let $\{u_k\}, k = 0, \ldots, N-1$ be the N frequency sampling points with sample spacing Δu. Then $u_k = k\Delta u$.
3. Let x_j^* be the return grid (the grid centers in our case) on which we want to obtain the values of $\gamma_t(x)$.
4. Now approximate the integral in (A.1) using the trapezoid rule, i.e.

$$\gamma_t(x) \approx \frac{1}{\pi}\sum_{k=0}^{N-1} \delta_k e^{-iu_k x}\tilde{\gamma}_t(u)\Delta u, \tag{A.2}$$

   where $\delta_k = \frac{1}{2}$ when $k = 0$ and 1 otherwise.
5. Then plugging the values of x_j^* into (A.2) we get

$$\gamma_t\left(x_j^*\right) \approx \frac{1}{\pi}\sum_{k=0}^{N-1} \delta_k e^{-iu_k x_j^*}\tilde{\gamma}_t(u_k)\Delta u. \tag{A.3}$$

Proof of Result 2: Under discretization, we replace $\boldsymbol{\lambda}_t = \{\boldsymbol{\lambda}_t(x), x \in [L, U]\}$ with $\boldsymbol{\lambda}_t$, a $B \times 1$ vector, following (4.5), similarly with $\boldsymbol{\gamma}_t$. Under the Gaussian process assumption for $\epsilon(x)$, we want to marginalize over $\boldsymbol{\lambda}$ when $\log\boldsymbol{\lambda} \sim GP(\log\boldsymbol{\gamma}, \Sigma)$. That is, we want to evaluate the following expression:

$$\begin{aligned}
[\mathbf{x}|\boldsymbol{\theta}] &= \int_0^{\infty}\left[\exp\left(-\sum_{i=1}^{B}\lambda(b)l\right)\prod_{i=1}^{B}[\lambda(b)]^{n_b}\right] \\
&\quad\times \frac{1}{(2\pi)^{B/2}|\Sigma|^{\frac{1}{2}}\prod_{i=1}^{B}\lambda(b)}\exp\left(-\frac{1}{2}(\log\boldsymbol{\lambda}-\log\boldsymbol{\gamma})'\Sigma^{-1}(\log\boldsymbol{\lambda}-\log\boldsymbol{\gamma})\right)d\boldsymbol{\lambda} \\
&= \exp\left[-\frac{1}{2}\left((\log\boldsymbol{\gamma})'\Sigma^{-1}(\log\boldsymbol{\gamma}) - \boldsymbol{\gamma}^{*\prime}\Sigma^{-1}\boldsymbol{\gamma}^*\right)\right] \\
&\quad\times \int_0^{\infty}\exp\left(-\sum_{b=1}^{B}e^{l\eta_i}\right)\exp\left(-\frac{1}{2}\left(\boldsymbol{\eta}-\boldsymbol{\gamma}^*\right)'\Sigma^{-1}\left(\boldsymbol{\eta}-\boldsymbol{\gamma}^*\right)\right)d\boldsymbol{\eta},
\end{aligned} \tag{A.4}$$

where $\boldsymbol{\eta} = \log \boldsymbol{\lambda}$, $\boldsymbol{\gamma}^* = \log \boldsymbol{\gamma} + \Sigma \times \mathbf{n}$, $\mathbf{n} = [n_1, \ldots, n_B]'$ and l is the length of the grid cell. Then note that the integral in (A.4) is actually $E_{\boldsymbol{\eta}}(e^{-\sum_{i=1}^{B} l e^{\eta_i}})$, where $\boldsymbol{\eta} \sim N(\boldsymbol{\gamma}^*, \Sigma)$. Then, following Harding and Hausman (2007), we write

$$E_{\boldsymbol{\eta}}\big(e^{-\sum_{i=1}^{B} l e^{\eta_i}}\big) = (2\pi)^{-B/2} |\Sigma|^{-\frac{1}{2}} \int \exp\big(-g(\boldsymbol{\eta})\big)\, d\boldsymbol{\eta},$$

where

$$\begin{aligned} g(\boldsymbol{\eta}) &= \frac{1}{2}\big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' \Sigma^{-1} \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big) + \sum_{i=1}^{B} l e^{\eta_i} \\ &= g_1(\boldsymbol{\eta}) + g_2(\boldsymbol{\eta}). \end{aligned} \tag{A.5}$$

Now expanding $g(\boldsymbol{\eta})$ about a point $\boldsymbol{\eta}^*$ using Taylor series we get

$$g(\boldsymbol{\eta}) \approx g\big(\boldsymbol{\eta}^*\big) + \big(\boldsymbol{\eta} - \boldsymbol{\eta}^*\big) \frac{\delta}{\delta \boldsymbol{\eta}} g(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} + \big(\boldsymbol{\eta} - \boldsymbol{\eta}^*\big)' \frac{\delta^2}{\delta \boldsymbol{\eta} \delta \boldsymbol{\eta}'} g(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} \big(\boldsymbol{\eta} - \boldsymbol{\eta}^*\big), \tag{A.6}$$

where $\boldsymbol{\eta}^*$ is the solution of the equation $\frac{\delta}{\delta \boldsymbol{\eta}} g(\boldsymbol{\eta}) = 0$, i.e.

$$\big(\boldsymbol{\eta}^* - \boldsymbol{\gamma}^*\big)' \Sigma^{-1} + \big[e^{l\eta_1^*}, \ldots, e^{l\eta_B^*}\big] = 0. \tag{A.7}$$

Equation (A.7) is a system of B non-linear equations in η and is difficult to solve simultaneously when B (number of grid cells) increases. So we expand $g_2(\boldsymbol{\eta})$ about $\boldsymbol{\gamma}^*$ using Taylor series. Then

$$g_2(\boldsymbol{\eta}) \approx g_2\big(\boldsymbol{\gamma}^*\big) + \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' \frac{\delta}{\delta \boldsymbol{\eta}} g_2(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\gamma}^*} + \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' \frac{\delta^2}{\delta \boldsymbol{\eta} \delta \boldsymbol{\eta}'} g_2(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\gamma}^*} \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big). \tag{A.8}$$

Now

$$J_2\big(\boldsymbol{\gamma}^*\big) = \frac{\delta}{\delta \boldsymbol{\eta}} g_2(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\gamma}^*} = l \times \big[e^{l\gamma_1^*}, \ldots, e^{l\gamma_B^*}\big]', \tag{A.9}$$

and

$$\Sigma_2\big(\boldsymbol{\gamma}^*\big) = \frac{\delta^2}{\delta \boldsymbol{\eta} \delta \boldsymbol{\eta}'} g_2(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\gamma}^*} = l \times \operatorname{diag}\big(l e^{l\gamma_1^*}, \ldots, l e^{l\gamma_B^*}\big). \tag{A.10}$$

Using (A.8), (A.9) and (A.10) in (A.5) we get

$$g(\boldsymbol{\eta}) \approx \frac{1}{2}\big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' \Sigma^{-1} \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big) + g_2\big(\boldsymbol{\gamma}^*\big) + \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' J_2\big(\boldsymbol{\gamma}^*\big) + \big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' \Sigma_2\big(\boldsymbol{\gamma}^*\big)\big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big). \tag{A.11}$$

Further expanding $g(\boldsymbol{\eta})$ is (A.11) about $\boldsymbol{\eta}^*$ using Taylor series we get

$$g(\boldsymbol{\eta}) \approx g\big(\boldsymbol{\eta}^*\big) + \big(\boldsymbol{\eta} - \boldsymbol{\eta}^*\big)' \frac{\delta}{\delta \boldsymbol{\eta}} g(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} + \big(\boldsymbol{\eta} - \boldsymbol{\eta}^*\big)' \frac{\delta^2}{\delta \boldsymbol{\eta} \delta \boldsymbol{\eta}'} g(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} \big(\boldsymbol{\eta} - \boldsymbol{\eta}^*\big), \tag{A.12}$$

where $\boldsymbol{\eta}^*$ is the solution of the equation $\frac{\delta}{\delta \boldsymbol{\eta}} g(\boldsymbol{\eta}) = 0$. Then differentiating $g(\boldsymbol{\eta})$ as obtained in (A.11) w.r.t. $\boldsymbol{\eta}$ we get

$$\begin{aligned} &\big(\boldsymbol{\eta} - \boldsymbol{\gamma}^*\big)' \big[\Sigma^{-1} + \Sigma_2\big(\boldsymbol{\gamma}^*\big)\big] + J_2\big(\boldsymbol{\gamma}^*\big) = 0 \\ &\Rightarrow \quad \boldsymbol{\eta}^* = \boldsymbol{\gamma}^* - \big[\Sigma^{-1} + \Sigma_2\big(\boldsymbol{\gamma}^*\big)\big]^{-1} J_2\big(\boldsymbol{\gamma}^*\big)'. \end{aligned} \tag{A.13}$$

Further differentiating w.r.t. $\boldsymbol{\eta}'$ we get

$$\Sigma^{*-1} = \frac{\delta^2}{\delta\boldsymbol{\eta}\delta\boldsymbol{\eta}'} g(\boldsymbol{\eta})\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}^*} = \left[\Sigma^{-1} + \Sigma_2(\boldsymbol{\gamma}^*)\right].$$

Then substituting Σ^{*-1} in (A.12) and completing the Gaussian integral we get

$$E_{\boldsymbol{\eta}}\left(e^{-\sum_{i=1}^{B} l e^{\eta_i}}\right) \approx \left[\frac{|\Sigma^*|}{|\Sigma|}\right]^{\frac{1}{2}} \exp\left(-g(\boldsymbol{\eta}^*)\right)$$

$$= \left[\frac{|\Sigma^*|}{|\Sigma|}\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\eta}^* - \boldsymbol{\gamma}^*)'\Sigma^{*-1}(\boldsymbol{\eta}^* - \boldsymbol{\gamma}^*) - \sum_{i=1}^{b} e^{\eta_i^*}\right] \text{(using (A.5))},$$

where $\boldsymbol{\eta}^*$ are obtained in closed form in (A.13). □

REFERENCES

Adler, P. B., Ellner, S. P., and Levine, J. M. (2010), "Coexistence of Perennial Plants: An Embarrassment of Niches," *Ecology Letters*, 13, 1019–1029.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman & Hall/CRC Press.

Banks, H. T., Kareiva, P. M., and Murphy, K. A. (1987), "Parameter Estimation Techniques for Interaction and Redistribution Models: A Predator-Prey Example," *Oecologia*, 74, 356–362.

Banks, H. T., Botsford, L. W., Kappell, F., and Wang, C. (1991), "Estimation of Growth and Survival in Size-Structured Cohort Data—An Application to Larval Striped Bass (*Morone Saxatilis*)," *Journal of Mathematical Biology*, 30, 125–150.

Caswell, H. (2001), *Matrix Population Models: Construction, Analysis and Interpretation* (2nd ed.), Sunderland: Sinauer.

——— (2008), "Perturbation Analysis of Nonlinear Matrix Population Models," *Demographic Research*, 18, 59–116.

Chakraborty, A., Gelfand, A. E., Wilson, A. M., Latimer, A. M., and Silander, J. A. (2011), "Point Pattern Modeling for Degraded Presence-Only Data Over Large Regions," *Journal of Royal Statistical Society, Series C*, 60 (5), 757–776.

Clark, J. S., Bell, D. M., Hersh, M. H., and Nichols, L. (2011a), "Climate Change Vulnerability of Forest Biodiversity: Climate and Resource Tracking of Demographic Rates," *Global Change Biology*, 17, 1834–1849.

Clark, J. S., Bell, D. M., Hersh, M. H., Kwit, M., Moran, E., Salk, C., Stine, A., Valle, D., and Zhu, K. (2011b), "Individual-Scale Variation, Species-Scale Differences: Inference Needed to Understand Diversity." *Ecology Letters*, 14, 1273–1287

Clark, J. S., Bell, D. M., Dietze, M., Hersh, M., Ibáñez, I., LaDeau, S., McMahon, S., Metcalf, J., Moran, E., Pangle, L., and Wolosin, M. (2010a), "Models for Demography of Plant Populations," in *The Oxford Handbook of Applied Bayesian Analysis*, eds. T. O'Hagan and M. West, London: Oxford University Press, pp. 431–481.

Clark, J. S., Bell, D., Chu, C., Courbaud, B., Dietze, M., Hersh, M., HilleRisLambers, J., Ibáñez, I., LaDeau, S., McMahon, S., Metcalf, J., Mohan, J., Moran, E., Pangle, L., Pearson, S., Salk, C., Shen, Z., Valle, D., and Wyckoff, P. (2010b), "High-Dimensional Coexistence Based on Individual Variation: A Synthesis of Evidence," *Ecological Monographs*, 80, 569–608.

Dahlgren, J. P., Garcia, M. B., and Ehrlén, J. (2011), "Nonlinear Relationships Between Vital Rates and State Variables in Demographic Models," *Ecology*. doi:*10.1890/10-1184.1*.

Dalgleish, H. J., Koons, D. N., Hooten, M. B., Moffet, C. A., and Adler, P. B. (2011), "Climate Influences the Demography of Three Dominant Sagebrush Steppe Plants," *Ecology*, 92, 75–85.

Davis, P. J., and Rabinowitz, P. (1984), *Methods of Numerical Integration* (2nd ed.), New York: Academic Press.

Dennis, B., Desharnais, R. A., Cushing, J. M., and Costantino, R. F. (1997), "Transitions in Population Dynamics: Equilibria to Periodic Cycles to Aperiodic Cycles," *Journal of Animal Ecology*, 66, 704–729.

Dennis, B., Desharnais, R. A., Cushing, J. M., and Costantino, R. F. (1995), "Nonlinear Demographic Dynamics: Mathematical Models, Statistical Methods, and Biological Experiments," *Ecological Monographs*, 65, 261–281.

Diggle, P. J. (2003), *Statistical Analysis of Spatial Point Patterns* (2nd ed.), London: Arnold.

Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000), "Size-Specific Sensitivity: Applying a New Structured Population Model," *Ecology*, 81, 694–708.

Ellner, S. P., and Rees, M. (2006), "Integral Projection Models for Species With Complex Demography," *The American Naturalist*, 167, 410–428.

——— (2007), "Stochastic Stable Population Growth in Integral Projection Models: Theory and Application," *Journal of Mathematical Biology*, 54, 227–256.

Harding, M. C., and Hausman, J. (2007), "Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Non-linear Least Squares," *International Economic Review*, 48, 1311–1328.

Hooten, M. B., Wikle, C. K., Dorazio, R. M., and Royle, J. A. (2007), "Hierarchical Spatio-Temporal Matrix Models for Characterizing Invasions," *Biometrics*, 63, 558–567.

Keyfitz, N., and Caswell, H. (2005), *Applied Mathematical Demography* (3rd ed.), New York: Springer.

Kot, M., Lewis, M., and van den Driessche, P. (1996), "Dispersal Data and the Spread of Invading Organisms," *Ecology*, 77, 2027–2042.

Rees, M., and Ellner, S. P. (2009), "Integral Projection Models for Populations in Temporally Varying Environments," *Ecological Monographs*, 79, 575–594.

Ricker, W. E. (1954), "Stock and Recruitment," *Journal of the Fisheries Research Board of Canada*, 11, 559–623.

Tierney, L., Kass, R. E., and Kadane, J. B. (1989a), "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *Journal of the American Statistical Association*, 84, 710–716.

——— (1989b), "Approximate Marginal Densities of Nonlinear Functions," *Biometrika*, 76, 425–433.

Tuljapurkar, S. (1990), *Population Dynamics in Variable Environments*, London: Springer.

Tuljapurkar, S., and Caswell, H. (eds.) (1997), *Structured-Population Models in Marine, Terrestrial, and Freshwater Systems*, New York: Chapman & Hall.

Wakefield, J. (2009), "Multi-Level Modelling, the Ecologic Fallacy, and Hybrid Study Designs," *International Journal of Epidemiology*, 38, 330–336.

Wikle, C. K. (2002), "A Kernel-Based Spectral Model for Non-Gaussian Spatio-Temporal Processes," *Statistical Modelling: An International Journal*, 2, 299–314.

Wikle, C. K., and Hooten, M. B. (2010), "A General Science-Based Framework for Nonlinear Spatio-Temporal Dynamical Models," *Test*, 19, 417–451.

Xu, K., Wikle, C. K., and Fox, N. (2005), "A Kernel Based Spatio Temporal Dynamical Model for Nowcasting Radar Precipitation," *Journal of the American Statistical Association*, 100, 1133–1144.

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261.

Zuidema, P. A., Jongejans, E., Chien, P. D., During, H. J., and Schieving, F. (2010), "Integral Projection Models for Trees: A New Parameterization Method and a Validation of Model Output," *Journal of Ecology*, 98, 345–355.