Lotka-Volterra competition models for sessile organisms

Matthew Spencer, Jason Tanner

May 2, 2007

1 Introduction

Markov models have been applied to data on a wide range of sessile communities, containing organisms such as trees, mussels and corals (Usher, 1979; Tanner et al., 1994; Wootton, 2001a; Hill et al., 2004), with the aim of understanding the structure and function of these communities. The popularity of Markov models stems from their relatively simple structure, and the ease with which they can be parameterized from data obtained by repeat surveys of permanent quadrats, which are a mainstay in marine ecology. There are also a wide range of tools that have been developed for analysis of Markov models of population-level data (Caswell, 2001), which can equally be applied to community models. While Markov models are very simple, in at least some cases they have proven to have surprisingly good predictive ability (Wootton, 2004).

In Markov models, a fixed point in space can be in one of a finite set of possible states (e.g. species or groups of species) at any given time, and the probabilities of future states depend only on the current state. In most cases, Markov models of communities are formulated in discrete time, with model time intervals based on the census intervals in the observed data. The relevant data are the frequencies of transitions from each state to each other state over a given time interval, and the parameters are the probabilities of these transitions.

Several elaborations of the basic first order discrete time Markov models generally used in ecology have been proposed and demonstrated. These include second-order models, where transitions depend on the state over the previous two time intervals, and semi-Markov models, where transitions depend on the length of time that a point has remained in its current state (Tanner et al., 1996), as well as a continuous time version (Spencer and Susko, 2005). It has often been suggested that the probabilities of transitions between states in Markov models of sessile communities might depend on densities (Usher, 1979; Tanner et al., 1994, 1996; Hill et al., 2002). However, there have been few attempts to incorporate density-dependence into these models. Acevedo (1981) studied the effects of density dependence on simple models of forest dynamics, while Caswell and Cohen have developed a number of nonlinear metapopulation models for two-species competition (e.g. Caswell and Cohen, 1995). More recently,

J. E. Tanner et al. (in review) have examined a more comprehensive density-dependent model of coral community dynamics, and found that it increased coral cover at equilibrium, compared to a model without density dependence.

All of these density-dependent models are formulated in discrete time. However, if transitions from one state to another may happen at any time, the transition probabilities over a finite time interval are the net outcome of all possible sequences of events in that interval. In most cases, a change in the rate of any single transition will then affect all transition probabilities (Spencer, 2006). We would therefore expect every transition probability to be affected by the abundance of every state, which leads to complicated models unless the number of states is very small. It is simpler to construct continuous-time density-dependent models if the biological interactions occur in continuous time. Here, we develop a continuous-time density-dependent model, based on a simple probabilistic view of interspecific interactions among sessile organisms. It turns out that this is a Lotka-Volterra competition model. Using maximum likelihood methods, we compare the fit of continuous-time models with and without density dependence to time series from a coral reef. In addition, we compare these models to time-averaged and saturated discrete-time models. We show that the density-dependent model performs much better than all but the saturated model. Furthermore, this improved performance is achieved by changing the form of the model, not by adding more parameters, in contrast to the saturated model, which requires a separate parameter for each transition in each time interval.

2 The model

2.1 Assumptions

We describe most of the features of the model in terms of colonial organisms such as corals, but the same approach can be applied to other kinds of sessile organisms such as trees. We make several important assumptions:

1. That there is a fixed and finite number of possible states for a point in space. Let s be the number of such states. One of these, e, is empty space, and the others may be either species or groups of species that we choose not to distinguish (either because this is too difficult or because they are of secondary interest).

2. That the rates of transitions between states depend only on current states, not on past states (the Markov assumption). This is not strictly true. For example: in some species, larger colonies are more likely to be dislodged by storms (Tanner et al., 1996; Madin and Connolly, 2006); reproduction depends on colony size in many corals (Harrison and Wallace, 1990); and competition between adjacent colonies may be size-dependent (Lang and Chornesky, 1990). Nevertheless, including historical effects in an empirical model of a reef system had little effect on community dynamics (Tanner

et al., 1996), so violations of the Markov assumption may not be very important.

3. That the rate at which transitions occur from state j into some non-empty state i depends on the availability of propagules or colonies of i to colonize or overgrow j, and that this availability depends on the proportion of i in the system. This assumption distinguishes our model from the usual homogeneous Markov chain, in which the rate of transition from j to i is a constant, independent of the proportion of i. We describe later how this assumption can be tested.

4. That the rate of clearance of points occupied by some species j is independent of the proportion of empty space in the system. Clearance might occur by external disturbances or because colonies of j die, and results in a transition to empty space. We assume that organisms do not interact such that a colony of i kills a colony of j but does not occupy the resulting space. Such interactions are biologically plausible, for example by allelopathy, but require more complex models.

We also assume that interaction coefficients are constant over time, that the system is of infinite extent and that local spatial effects are unimportant. These latter assumptions are not likely to be true. However, our aim is to produce a simple model which can easily be tested using field data, and to evaluate the influence of state frequencies on interaction rates. We formulate the resulting model as a mean-field system of constant-coefficient nonlinear differential equations. Transitions between states may occur at any time. There is no reason to assume organisms only interact at fixed moments in time, unlike models of organisms with annual lifecycles, where discrete time is a natural choice. However, the properties of the system are likely to be sampled at discrete points in space and time. We therefore base our likelihood function on discrete sampling.

2.2 Derivation

Let Ω be the extent of the system, which we assume here is infinite. Consider a point $w \in \Omega$ whose state is X_w. Let $v \in \Omega$ be another point in the system, and let $\lambda_{ij}(v, w)$ (dimensions T^{-1}) be the finite rate at which dispersal or growth from the colony at v causes transitions from state j to state i at w, if $X_v = i$, $i \neq e$, and $X_w = j$. We assume that $\lambda_{ij}(v, w)$ is constant over time and does not depend on the states of any other points. Integrating over all pairs of points in the system, the total rate at which transitions occur from j to i is

$$\mu_{ij} \equiv \int_{w \in \Omega} \int_{v \in \Omega} \lambda_{ij}(v, w) I\{v, i\} I\{w, j\} dv dw \tag{1}$$

where $I\{v, i\}$ is an indicator variable with value 1 if $X_v = i$ and 0 otherwise. We assume that this integral converges to a finite value, which will be true if the probability of dispersal or growth from v to w decays sufficiently fast with distance. However, we have no information on the relevant dispersal and growth

distributions, so we make a mean-field approximation. Let the expectations over Ω of $\lambda_{ij}(v,w)$, $I\{v,i\}$ and $I\{w,j\}$ be a_{ij} (dimensions T^{-1}), x_i (dimensionless) and x_j (dimensionless) respectively, at a given moment in time. We assume that the only relevant information about a pair of points is their states, ignoring any spatial effects. Then

$$\mu_{ij} \approx a_{ij} x_i x_j \tag{2}$$

If there are spatial effects, we can justify Equation 2 as the first term in a Taylor series approximation.

The proportion of points in state j at any time is x_j. We refer to $a_{ij}x_i$ as the instantaneous rate of transitions per unit time from j to non-empty state i in this model, per unit frequency of the source state j. The relationship between instantaneous rates and transition probabilities is the same as that between the instantaneous growth rate of a population (defined by a differential equation) and the ratio of population sizes at two times $t+1$ and t (defined by a difference equation). As we will show later, the rates of change can be integrated numerically to obtain the probabilities of transitions from state j to state i over a finite interval of time. We refer to a_{ij} as an interaction coefficient.

For a point w in the system that is in a non-empty state j, we assume that the rate of transitions $\lambda_{ej}(w)$ (dimensions T^{-1}) to the empty state e is independent of the states of all other points. Then under a similar mean-field assumption the rate of transitions from j to e is

$$\mu_{ej} = \int_{w\in\Omega} \lambda_{ej}(w) I\{w,j\} dw \approx a_{ej} x_j \tag{3}$$

We refer to a_{ej} as the rate of transitions from non-empty state j to empty state e in this model. Again, this rate is per unit frequency of the source state j.

We can now write an equation for the rate of change of frequency of each state, by summing the loss and gain terms over all destination and source states:

$$\frac{dx_i}{dt} = \begin{cases} -\left(a_{ei} + \sum_{k\neq e,i} a_{ki} x_k\right) x_i + \sum_{k\neq i} a_{ik} x_i x_k, & i \neq e \\ -\sum_{k\neq e} a_{ke} x_k x_e + \sum_{k\neq i} a_{ek} x_k, & i = e \end{cases} \tag{4}$$

For simplicity, we do not discuss facilitation in detail here, but it could be included in such models. For example, if species i is better able to colonize empty space when the abundance of another species j is greater, we might include terms like $a_{ie,j} x_i x_e x_j$, where $a_{ie,j}$ (dimensions T^{-1}) is the coefficient of colonization of empty space by i per unit frequency of j.

We can rewrite Equation 4 in matrix form. Let $\mathbf{A}$ be a matrix whose off-diagonal elements are the interaction coefficients a_{ij} and whose diagonal elements are zero. Let $\mathbf{X}$ be a diagonal matrix with entries x_i if $i \neq e$, and 1 if $i = e$. Let $\mathbf{C}$ be a diagonal matrix of column sums of $\mathbf{XA}$. Let $\mathbf{x}$ be a column vector of probabilities of each state. Then

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= (\mathbf{XA} - \mathbf{C})\mathbf{x} \\ &= \mathbf{R}(\mathbf{x})\mathbf{x} \end{aligned} \tag{5}$$

where $\mathbf{R}(\mathbf{x})$ is a density-dependent rate matrix.

3 Relationships to other models

In this section, we show how the model of Section 2 is related to two well-known ecological models. First, it is a Lotka-Volterra competition model. Second, it is indistinguishable from a homogeneous continuous-time linear Markov model (or its discrete-time equivalent) if it is at equilibrium, but will behave differently away from equilibrium and will respond differently to changes in parameters.

3.1 Lotka-Volterra competition model

The general Lotka-Volterra competition model is

$$\frac{dx_i}{dt} = r_i x_i - \frac{r_i}{K_i} x_i^2 - \sum_{k \neq i} \frac{r_i}{K_i} \alpha_{ik} x_i x_k \tag{6}$$

(MacArthur and Levins, 1967), where r_i (dimensions T^{-1}) is a per-capita population growth rate, K_i (dimensionless) is a carrying capacity, and α_{ik} (dimensionless) is an interspecific competition coefficient measuring the effect of species k on the growth rate of species i.

Because every point in the system is in one of the possible states, the proportion of points that are empty can be written as $x_e = 1 - \sum_{k \neq e} x_k$. Substituting this into Equation 4 with $i \neq e$ and rearranging, we obtain

$$\frac{dx_i}{dt} = (a_{ie} - a_{ei}) x_i - a_{ie} x_i^2 - \sum_{k \neq e,i} (a_{ki} + a_{ie} - a_{ik}) x_i x_k \tag{7}$$

which is identical to Equation 6 with $r_i = a_{ie} - a_{ei}$, $K_i = (a_{ie} - a_{ei})/a_{ie}$, and $\alpha_{ik} = (a_{ki} + a_{ie} - a_{ik})/a_{ie}$. We therefore refer to the model of Section 2 as the LV model from now on.

The Lotka-Volterra competition model can also arise from completely different mechanistic assumptions, or simply as an approximation to a more complex model close to equilibrium (Schoener, 1986).

3.2 Linear Markov models

If Equation 5 is at equilibrium, $\mathbf{R}(\mathbf{x})$ does not vary over time. It is therefore indistinguishable at equilibrium from the homogeneous continuous-time linear Markov model

$$\frac{d\mathbf{x}}{dt} = \mathbf{Q}\mathbf{x} \tag{8}$$

where $\mathbf{Q}$ is a matrix whose off-diagonal elements q_{ij} (dimensions T^{-1}) are non-negative instantaneous transition rates, and whose diagonal elements q_{jj} are -1 times the column sums of off-diagonal elements. We refer to this model from now on as the linear model. Elsewhere (Spencer and Susko, 2005), we discuss

the relationship between this linear model and the usual discrete-time Markov models for communities of sessile organisms. Briefly, the usual formulation of a discrete-time Markov model is

$$\mathbf{x}(T+t) = \mathbf{P}(t)\mathbf{x}(T) \tag{9}$$

where $\mathbf{x}(T)$ is a vector of state probabilities at time T, and $\mathbf{P}(t)$ is a transition probability matrix whose ijth entry $p_{ij}(t)$ is the conditional probability of observing state i at time $T+t$ given that we observed state j at time T. If there is a homogeneous continuous-time process $\mathbf{Q}$, then

$$\begin{aligned} \mathbf{P}(t) &= \sum_{m=0}^{\infty} \frac{(\mathbf{Q}t)^m}{m!} \\ &= e^{\mathbf{Q}t} \end{aligned} \tag{10}$$

where $e^{\mathbf{Q}t}$ is a matrix exponential. $\mathbf{P}(t)$ is a stochastic matrix, and its largest eigenvalue is 1. Most models of this kind have a globally stable stationary distribution (Hill et al., 2004).

In many ecological analyses, the $\mathbf{P}$ matrix is estimated by recording the identities of species at fixed points in space at a series of evenly-spaced time intervals, and aggregating the transition counts over space and time (e.g. Tanner et al., 1994; Wootton, 2001c; Hill et al., 2004). To do so, we must assume either that the $\mathbf{P}$ matrix is independent of state frequencies, or that the frequencies are close to equilibrium.

4 Likelihood function

One appealing feature of models for sessile organisms observed at discrete time intervals is that we can easily derive the likelihood of a model given the data. We can then make formal comparisons between different models. We first derive the likelihood function, then calculate the transition probabilities between states under each model.

4.1 The product multinomial likelihood

Suppose we have a sequence of states $a_0, a_1, \ldots a_k$ at a point in space observed at times $t_0, t_1, \ldots t_k$, where the time intervals are not necessarily equal. Under the Markov assumption, the probability of this sequence is

$$\begin{aligned} P(a_0, a_1, \ldots a_k) &= P(a_k|a_{k-1})P(a_{k-1}|a_{k-2}) \ldots P(a_1|a_0)P(a_0) \\ &= P(a_0) \prod_{m=1}^{k} P(a_m|a_{m-1}) \end{aligned} \tag{11}$$

where $P(a_m|a_{m-1})$ is the probability of observing state a_m at time t_m given state a_{m-1} at time t_{m-1} and $P(a_0)$ is the probability of the initial state.

If we have a sample of sequences from a set of v independent and identically distributed (iid) points, then the likelihood L for the sequences at all the points is the product multinomial

$$\begin{aligned} L &= \prod_{h=1}^{v} P(a_{0,h}) \prod_{m=1}^{k} P(a_{m,h}|a_{m-1,h}) \\ &= [\prod_{j} p_j(0)^{n_j(0)}] \prod_{m=1}^{k} \prod_{ij} p_{ij}(m, m-1)^{n_{ij}(m,m-1)} \end{aligned} \tag{12}$$

where $a_{m,h}$ is the state at point h at time t_m, $p_j(0)$ is the probability of state j at time 0, $p_{ij}(m, m-1)$ is the probability of state i at time t_m given state j at time t_{m-1}, $n_j(0)$ is the number of points in state j at time 0 and $n_{ij}(m, m-1)$ is the number of points in state j at time t_{m-1} and state i at time t_m. The product $\prod_j$ is over all states and the product $\prod_{ij}$ is over all combinations of states. In practice, it is easier to work with the log likelihood

$$l = \sum_{j} n_j(0) \log p_j(0) + \sum_{m=1}^{k} \sum_{ij} n_{ij}(m, m-1) \log p_{ij}(m, m-1) \tag{13}$$

In a homogeneous Markov model, initial states do not affect transition probabilities over subsequent time intervals, so the initial state is an ancillary statistic if the interaction rates or probabilities are the parameters of interest and we do not assume the process is at equilibrium. It is then usual to consider only the likelihood of the sequence conditional on the initial state (Lehmann, 1986, section 10.2). However, in an LV model, initial states do affect subsequent transition probabilities, so we will use full likelihoods in all cases.

Equation 13 assumes that the sample points were independent. This implies the assumption that each sampled point has a negligible effect on any other sample points and on the dynamics of the entire system. This may be reasonable if the system is large and either interactions are not strongly local or sample points are far apart. Even if the iid assumption is incorrect, the model may still be useful. Since the model describes the behaviour of a single point in space, parameter estimates from the marginal distribution of states at a point will be asymptotically correct. However, if there are dependencies, it will be as if there were fewer independent observations than sample points, so hypothesis tests should be interpreted cautiously. This could be addressed in future by working with spatially resolved data.

4.2 Transition probabilities in the LV model

We now need to calculate $p_{ij}(m, m-1)$ for the model specified by Equation 4. For simplicity, we will set $t_{m-1} = 0$ and write $p_{ij}(t)$ for the probability that a sample point in state j at time 0 is in state i at time $t \geq 0$. The initial condition is $p_{ij}(0) = 0$ if $i \neq j$ and $p_{jj}(0) = 1$, because at time 0 there is no possibility of any change of state. The probability p_{ij} will undergo losses due to transitions

to states other than i, and gains due to transitions into state i from points now in some other state k that were in state j at time 0. The rate of change of p_{ij} thus has the same form as the rate of change of x_i (Equation 4):

$$\frac{dp_{ij}}{dt} = \begin{cases} -\left(a_{ei} + \sum_{k \neq e,i} a_{ki} x_k\right) p_{ij} + \sum_{k \neq i} a_{ik} x_i p_{kj} & i \neq e \\ -\sum_{k \neq e} a_{ke} x_k p_{ej} + \sum_{k \neq e} a_{ek} p_{kj}, & i = e \end{cases} \tag{14}$$

Let $\mathbf{p}_j$ be a column vector of probabilities of each state conditional on being in state j at time 0. Then we can write Equation 14 in matrix form by substituting $\mathbf{p}_j$ for $\mathbf{x}$ (the unconditional probabilities of each state) in Equation 5:

$$\begin{aligned} \frac{d\mathbf{p}_j}{dt} &= (\mathbf{XA} - \mathbf{C})\mathbf{p}_j \\ &= \mathbf{R}(\mathbf{x})\mathbf{p}_j \end{aligned} \tag{15}$$

We can integrate Equations 5 and 15 numerically to get the transition probabilities for any time interval, and insert these into Equation 13 to get the log likelihood for any given parameters. This model has $s^2 - 1$ parameters: $s(s-1)$ interaction rates and $s - 1$ initial state probabilities. The initial probability of the last state is determined by the sum of the probabilities of the other states, so there are only $s - 1$ independent probabilities.

4.3 Transition probabilities in the linear model

For a homogeneous continuous-time linear Markov model (Equation 8), the transition probabilities are given by the exponential of the $\mathbf{Q}$ matrix. Again, this model has $s^2 - 1$ parameters (including the initial state probabilities, although these do not affect transition probabilities).

4.4 Transition probabilities in the saturated discrete-time Markov model

The best possible model fits a separate transition probability matrix to each time interval. The maximum likelihood estimates of transition probabilities are given by

$$\hat{p}_{ij}(m, m-1) = \frac{n_{ij}(m, m-1)}{\sum_k n_{kj}(m, m-1)} \tag{16}$$

(Caswell, 2001, page 135). This model has $(ks+1)(s-1)$ parameters including initial state probabilities. There is one parameter for every category of observation, so this is a saturated model.

4.5 Transition probabilities in the time-averaged discrete Markov model

If samples are taken at equal intervals (or under the hypothesis that transition probabilities do not depend on the time interval), we could force the transition

probabilities to be the same for all intervals. This gives time-averaged maximum likelihood estimates

$$\hat{p}_{ij} = \frac{\sum_m n_{ij}(m, m-1)}{\sum_m \sum_k n_{kj}(m, m-1)} \tag{17}$$

This model has $s^2 - 1$ parameters, including initial state probabilities.

5 Identifiability

If a model is going to help us understand the workings of a community, we have to be able to estimate its parameters. A model is identifiable if all its parameters can be estimated from the data. In other words, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \implies l(\boldsymbol{\theta}) \neq l(\boldsymbol{\theta}_0)$, for two parameter vectors $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a parameter vector at which the likelihood is maximized. In particular, a model will not be identifiable if some of its parameters are redundant, so that the model can be rewritten with a smaller number of parameters without changing the likelihood. For example, the linear model $\mathbf{y} = \theta_0 + (\theta_1 + \theta_2)\mathbf{x}$ with $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2]^{\mathrm{T}}$ is not identifiable, because we could obtain the same likelihood from $\boldsymbol{\theta} = [\theta_0, \theta_1 + \alpha, \theta_2 - \alpha]^{\mathrm{T}}$ for any α. Identifiability depends on the structure of the model, not just the number of parameters. For example, $\mathbf{y} = \theta_0 + \theta_1\mathbf{x} + \theta_2\mathbf{x}^2$ has the same number of parameters as the previous example, but may be identifiable, because θ_1 and θ_2 now affect the likelihood in different ways.

For more complex models, it is often not obvious whether there are redundant parameters. One way to determine this is to calculate the rank of the Jacobian matrix $\mathbf{D}$, where $d_{ij} = \frac{\partial \mu_j}{\partial \theta_i}$ and μ_j is the expected value of the jth class of observations. Each row of this matrix gives the effects of changing one parameter on all the expected values. The rank of a matrix is the number of linearly independent rows, and a matrix is of full rank if all its rows are linearly independent. If $\mathbf{D}$ is not of full rank, then there is a nonzero vector $\boldsymbol{\alpha}(\boldsymbol{\theta})$ such that $\boldsymbol{\alpha}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{D}(\boldsymbol{\theta}) = \mathbf{0}$. If we take $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, then $\nabla l\boldsymbol{\alpha}(\boldsymbol{\theta}) = 0$ (Catchpole and Morgan, 1997, theorem 2). In other words, moving in the direction given by $\boldsymbol{\alpha}(\boldsymbol{\theta})$ does not change the likelihood. Intuitively, this means that there is a ridge of parameter values all having the same likelihood, and the model is not identifiable. However, there are cases where a model is not identifiable even though the Jacobian is of full rank (Catchpole and Morgan, 1997).

We illustrate the relationship between the rank of the Jacobian and identifiability by returning to the linear models above. Consider the parameter-redundant case $\mathbf{y} = \theta_0 + (\theta_1 + \theta_2)\mathbf{x}$, treating $\mathbf{x} = [x_1 < x_2 < \ldots x_n]$ as fixed. The Jacobian is

$$\mathbf{D} = \begin{bmatrix} \partial y_1/\partial\theta_0 & \partial y_2/\partial\theta_0 & \ldots & \partial y_n/\partial\theta_0 \\ \partial y_1/\partial\theta_1 & \partial y_2/\partial\theta_1 & \ldots & \partial y_n/\partial\theta_1 \\ \partial y_1/\partial\theta_2 & \partial y_2/\partial\theta_2 & \ldots & \partial y_n/\partial\theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1 & x_2 & \ldots & x_n \end{bmatrix} \tag{18}$$

This has rank 2, because the second and third rows are identical. Solving $\boldsymbol{\alpha}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{D}(\boldsymbol{\theta}) = \mathbf{0}$ gives $\boldsymbol{\alpha}(\boldsymbol{\theta})^{\mathrm{T}} = [0, \alpha, -\alpha]$, as expected.

The time-averaged discrete and saturated models are known to be identifiable. For the LV and linear models, we do not have closed-form expressions for the Jacobian (or for the Fisher information matrix, which has the same rank as the Jacobian) so they must be evaluated numerically for particular parameter values, and we will not be able to prove that the models are always identifiable. In the absence of numerical errors, the rank of a matrix is equal to the number of non-zero singular values it possesses (Horn and Johnson, 1985, p. 414), so in practice we estimate the rank as the number of singular values greater than some small positive constant. We treated the observation times as fixed, and estimated the rank of the Jacobian at the estimated parameter values from the data sets described below for the LV and linear models. We did not find any problems with identifiability of the LV model. However, there were potential problems with identifiability of linear models for some estimated parameters, which we discuss below. Continuous-time linear Markov models are not always identifiable from discrete-time data (Singer and Spilerman, 1976). This does not affect comparisons between models, but may make it difficult to interpret parameter estimates from the linear models.

6 Estimation

Parameter estimation requires maximizing the log likelihood as a function of the parameters. For a discrete-time model, the maximum likelihood estimates of transition probabilities are given by equations 16 or 17. We do not have closed-form estimates for the other models, so we use numerical optimization as described in the Appendix.

7 Model selection

The LV, linear, and time-averaged discrete models (and all other possible models) are nested within the saturated model, so we can use likelihood ratio statistics to compare each to the saturated model (Hilborn and Mangel, 1997, pages 153-154). Asymptotically, the test statistic $2(l_{\text{saturated}} - l)$ has a $\chi^2_{\Delta p}$ distribution, where $l_{\text{saturated}}$ is the log likelihood of the saturated model, l is the log likelihood of the other model, and Δp is the difference in the number of parameters between the saturated model and the other model (Bickel and Doksum, 2001, section 6.3.1). However, our non-saturated models are not nested and all have the same number of parameters, so the preferred model is the one with the largest log likelihood. We can also compare all four models using Akaike's Information Criterion $\text{AIC}_k = -2l + 2p$, where p is the number of parameters for model k (Akaike, 1992; Bozdogan, 1987). The preferred model is the one with the smallest AIC (Hilborn and Mangel, 1997, pages 159-160). The relative likelihood of a model k can be asymptotically approximated by $l_k = \exp((\text{AIC}_0 - \text{AIC}_k)/2)$, where AIC_0 is the AIC of the best model (Burnham and Anderson, 2004). The Akaike weight $w_k = l_k / \sum_{j \in \mathcal{M}} l_j$ can be interpreted

as an estimate of the probability that model k is the best in the set $\mathcal{M}$ of models under consideration according to the AIC criterion (Burnham and Anderson, 2004), although this interpretation is not without controversy (Link and Barker, 2006).

8 Data

We fitted the models to data from a long-term study of coral community dynamics at Heron Island, Great Barrier Reef, Queensland, Australia (Connell et al., 1997, 2004). Data from this study have previously been analyzed using both discrete-time (Tanner et al., 1994, 1996; Hill et al., 2004) and continuous-time (Spencer and Susko, 2005; Spencer, 2006) Markov models, and are available on request from JT. Data were collected from photographs of fixed $1m^2$ quadrats taken at unequal intervals over 27 years, from 1962 to 1989. Grids of points were placed over the photographs and the species present at each point recorded as described in Tanner et al. (1994). 72 species of corals and 9 species of algae were observed in the quadrats. These were grouped into eight categories plus a free space state in previous analyses (Tanner et al., 1994, 1996). Free space was usually occupied by organisms such as crustose corraline and turfing algae, but is available for colonization by corals and macroalgae (J.E. Tanner et al., in review).

Initial analyses (Appendix) showed that the reliability of parameter estimation was improved by aggregating the four acroporid coral states into a single state, reducing the number of parameters from 80 to 35. Aggregation is reasonable because all the acroporids showed similar trends in frequency over time. Analyses of the unaggregated data led to qualitatively similar conclusions. Simulation studies (Appendix) showed that we could correctly identify the true model and obtain good parameter estimates in most cases. We report results from only one intertidal site, the Protected Crest. Data were also available for one other intertidal site (Exposed Crest) and one subtidal site (Exposed Pool). However, these had fewer sample points in time and/or space, and simulation studies like those described in the Appendix showed that there were too few data for reliable parameter estimation. There were small numbers of missing observations ($< 1\%$), which we ignore. We analyzed data from all 17 sample dates used by Tanner et al. (1994). There were at least 1249 points observed per time interval.

9 Results and discussion

9.1 Model selection

Table 1 shows the log likelihoods l, number of parameters p, and Akaike's Information Criterion for all the models. The saturated model is much better than the LV model, which in turn is much better than the linear model. The time-averaged discrete model is worst of all.

Likelihood ratio tests reject the LV, linear and time-averaged discrete models with $p < 1 \times 10^{-16}$. The small p-value is not surprising because there are 450 degrees of freedom. The Akaike weight of the saturated model is > 0.9999, indicating overwhelming support for this model compared to the others.

The saturated model is not of much biological interest. Although it is the best possible description of the data, it tells us nothing about mechanisms, and it cannot be used to predict future events. Its main value is to provide a standard against which other models are measured. Thus, even though our other models can be rejected as a complete explanation of the data, it is still worth comparing them to each other in order to choose the most promising framework for further development. Similar situations arise in modelling molecular evolution: early models could often be rejected (e.g. Huelsenbeck and Crandall, 1997, page 454), but have provided a basis for the development of more sophisticated models (Sullivan and Joyce, 2005, page 459).

The LV model has a much smaller AIC than the linear model. If the saturated model is excluded from the comparison, the Akaike weight of the LV model is > 0.9999. Thus, the LV model is much better than any other non-saturated model we considered. This indicates that transition probabilities are likely to depend on state frequencies, although comparison with the saturated model shows that other factors must also be important. Both continuous-time models are better than the time-averaged discrete model. It is not surprising that transition probabilities depend on the length of the time interval, although most previous models have not taken account of this (e.g. Tanner et al., 1994, 1996; Spencer and Susko, 2005). However, there are plausible situations in which simple continuous-time models would not work well, for example if temporal variability in environmental conditions mattered more than the length of the time interval.

Figure 1 shows the predicted and observed frequencies of each state at each sample date. Parameters are estimated from transition frequency data, but time series of abundance are a good visual representation of the behaviour of each model. Predicted frequencies are given by Equation 5 for the LV model, and Equation 8 for the linear model. For the time-averaged discrete model we generated expected frequencies using Equation 9, ignoring the variation in sample intervals. The LV model generates predicted frequencies that look much more like the observed data than those from the linear model. The time-averaged discrete model generates frequencies similar to the linear model.

In the linear model, the smallest singular value of the Jacobian was 1×10^{-9}, which may indicate potential identifiability problems. The largest transition rate was from algae to free space ($q_{63} = 147.53$), an order of magnitude larger than any other. The predicted proportion of algae is low and rapidly approaches an equilibrium. Small changes in other transition rates involving algae are unlikely to have much effect on this behaviour. There may not be very well-defined optimal parameter estimates for the linear model. However, since the linear model performs much worse than the LV model, these estimates are not of much interest. Although there were some large rates in the LV model, the rate matrix was not dominated by a single large rate and no singular values were

less than 1×10^{-5}. We can therefore be more confident that the parameter estimates for the LV model have biological meaning.

9.2 Interaction coefficient estimates in the LV model

Tables 2, 3 and 4 show the parameter estimates for all the non-saturated models. Here, we briefly discuss the biological significance of the interaction coefficient estimates in the LV model.

High coefficients for transitions into a state are not necessarily associated with high abundance. For example, there are high coefficients for transitions from acroporids, massive corals and free space to algae. However, there are also high coefficients for transitions from algae to pocilloporid corals and free space. Algae therefore show rapid turnover but do not become abundant (Figure 1). This is in accordance with the idea that algae are transient, fast-colonizing species on this reef (Connell, 1987).

There are a number of very low coefficients ($< 1 \times 10^{-9}$: 5/30 rates = 17%). In earlier analyses with four separate acroporid states, almost every possible transition occurred at some point during the observation period (Tanner et al., 1994), although a continuous-time linear model suggested that some transitions may only have occurred indirectly (Spencer and Susko, 2005). When acroporids were aggregated into a single state, the proportion of possible transitions that were never observed (7/36 = 19%) was not dramatically different from the proportion of very low coefficients in the LV model, although only three of these unobserved transitions also had coefficients less than 1×10^{-9} in the LV model. When some states are very transient, others persist for much longer, and the sampling intervals are moderately long, the pattern of transitions that are observed may not accurately reflect the events that actually occur (J.E. Tanner et al., in review).

10 Conclusions

For the data set we studied, the Lotka-Volterra model performed much better than two density-independent alternatives, even though all these non-saturated models had the same number of parameters. We also analyzed data from two other sites, the Exposed Crest and Exposed Pools, but do not report these results here. The other two sites had shorter time series and/or fewer points in space, and simulations showed that parameter estimation was less reliable at these sites than at the Protected Crest. Analyses of both these sites strongly favoured the LV model over the linear model, as at the Protected Crest. However, simulations showed that the frequency of wrongly selecting the LV model when the linear model was the true model was much higher than at the Protected Crest. As mentioned previously, a linear model close to equilibrium may be difficult to distinguish from an LV model close to equilibrium, and the linear models quickly approached equilibrium for parameters estimated from the field data. Although this may not be a problem for the field data, because the sys-

tems were not particularly close to equilibrium, we are reluctant to draw any strong conclusions from these other sites. This highlights the need for long time series as a foundation for statistical ecology. We hope that further studies on other long-term ecological data sets will lead to a more general understanding of the situations in which linear and LV models are suitable for communities of sessile organisms.

One of the other sites we analyzed, the Exposed Pools, was also strongly affected by storm damage on several occasions (Connell et al., 1997). The LV model was able to reproduce some of the observed large fluctuations in abundance of corals and free space following these storms. However, in reality these fluctuations were likely to have been a consequence of fluctuations in rates of transitions to free space, which were treated as constant in the model. Mortality rates are likely to vary considerably over time, to depend on the time since the last storm (because larger colonies may be more vulnerable to storm damage), and to covary among points in space and states in the system. It therefore seems unlikely that the estimated parameters from the LV model are biologically meaningful when occasional major storm damage is important. It should be possible to explicitly include storm damage in the models described here, by using data on storm intensities as a predictor of transitions to free space (Madin and Connolly, 2006).

Other factors may also be changing over time. At the Protected Crest site, acroporids were initially abundant but showed a gradual decline, free space was increasing over time, massive corals were moderately abundant and soft corals showed a rapid increase at the end of the time series (Figure 1). These trends may be partly due to upward growth of the reef and partial diversion of water flow (Connell et al., 2004).

Another possible improvement is making the model explicitly spatial. It is likely that transition rates at a point depend more strongly on the states of the immediate neighbours of the point than on the average state of the system. Although there are explicit spatial models of sessile communities (e.g. Wootton, 2001b; Langmead and Sheppard, 2004), little effort has yet been expended on their probabilistic foundations, fitting to time series, or performance relative to non-spatial models. There is ample scope for further development of testable stochastic models in this field.

Acknowledgments

This work was supported by a grant from the University of Liverpool Research Development Fund, project number 6119. We are particularly grateful to Joe Connell for initiating and conducting the field study, and making the data available, and Terry Hughes for his role in data collection and processing. We are also grateful to Steve Brooks and Damian Clancy for helpful suggestions, and to Mary Thorp and Cliff Addison for technical support.

References

Acevedo, M. F. (1981). On Horn's Markovian model of forest dynamics with particular reference to tropical forests. *Theoretical Population Biology*, 19(2):230–250.

Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics Volume I: Foundations and Basic Theory*, pages 599–624. Springer-Verlag, New York.

Bickel, P. J. and Doksum, K. A. (2001). *Mathematical statistics: basic ideas and selected topics*, volume 1. Prentice-Hall, New Jersey, second edition.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, 52(3):345–370.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2):261–304.

Caswell, H. (2001). *Matrix population models: construction, analysis, and interpretation*. Sinauer, Sunderland, MA, second edition.

Caswell, H. and Cohen, J. E. (1995). Red, white and blue: environmental variance spectra and coexistence in metapopulations. *Journal of Theoretical Biology*, 176:301–316.

Catchpole, E. A. and Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika*, 84(1):187–196.

Connell, J. H. (1987). Change and persistence in some marine communities. In Gray, A. J., Crawley, M. J., and Edwards, P. J., editors, *Colonization, Succession and Stability*, pages 339–352. Blackwell Scientific Publications, Oxford.

Connell, J. H., Hughes, T. P., and Wallace, C. C. (1997). A 30-year study of coral abundance, recruitment, and disturbance at several scales in space and time. *Ecological Monographs*, 67(4):461–488.

Connell, J. H., Hughes, T. P., Wallace, C. C., Tanner, J. E., Harms, K. E., and Kerr, A. M. (2004). A long-term study of competition and diversity of corals. *Ecological Monographs*, 74(2):179–210.

Harrison, P. L. and Wallace, C. C. (1990). Reproduction, dispersal and recruitment of scleractinian corals. In Dubinsky, Z., editor, *Ecosystems of the World 25: Coral Reefs*, pages 133–207. Elsevier, Amsterdam.

Hilborn, R. and Mangel, M. (1997). *The ecological detective*. Princeton University Press, Princeton, New Jersey.

Hill, M. F., Witman, J. D., and Caswell, H. (2002). Spatio-temporal variation in Markov chain models of subtidal community succession. *Ecology Letters*, 5:665–675.

Hill, M. F., Witman, J. D., and Caswell, H. (2004). Markov chain analysis of succession in a rocky subtidal community. *American Naturalist*, 164:E46–E61.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.

Huelsenbeck, J. P. and Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, 28:437–466.

Lang, J. C. and Chornesky, E. A. (1990). Competition between scleratinian reef coras — a review of mechanisms and effects. In Dubinsky, Z., editor, *Ecosystems of the World 25: Coral Reefs*, pages 209–252. Elsevier, Amsterdam.

Langmead, O. and Sheppard, C. (2004). Coral reef community dynamics and disturbance: a simulation model. *Ecological Modelling*, 175:271–290.

Lehmann, E. L. (1986). *Testing statistical hypotheses*. John Wiley and Sons, second edition. 600 pp.

Link, W. A. and Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology*, 87(10):2626–2635.

MacArthur, R. and Levins, R. (1967). The limiting similarity, covergence and divergence of coexisting species. *American Naturalist*, 101:377–385.

Madin, J. S. and Connolly, S. R. (2006). Ecological consequences of major hydrodynamic disturbances on coral reefs. *Nature*, 444:477–480.

Schoener, T. W. (1986). Mechanistic approaches to community ecology: a new reductionism? *American Zoologist*, 26:81–106.

Singer, B. and Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology*, 82(1):1–54.

Spencer, M. (2006). Sensitivity analysis of Markov models for communities of competing sessile organisms. *Journal of Animal Ecology*, 75:1024–1033.

Spencer, M. and Susko, E. (2005). Continuous-time Markov models for species interactions. *Ecology*, 86(12):3272–3278.

Sullivan, J. and Joyce, P. (2005). Model selection in phylogenetics. *Annual Review of Ecology and Systematics*, 36:445–466.

Tanner, J. E., Hughes, T. P., and Connell, J. H. (1994). Species coexistence, keystone species, and succession: a sensitivity analysis. *Ecology*, 75(8):2204–2219.

Tanner, J. E., Hughes, T. P., and Connell, J. H. (1996). The role of history in community dynamics: a modelling approach. *Ecology*, 77(1):108–117.

Usher, M. B. (1979). Markovian approaches to ecological succession. *Journal of Animal Ecology*, 48:413–426.

Wootton, J. T. (2001a). Causes of species diversity differences: a comparative analysis of Markov models. *Ecology Letters*, 4(1):46–56.

Wootton, J. T. (2001b). Local interactions predict large-scale pattern in empirically derived cellular automata. *Nature*, 413:841–844.

Wootton, J. T. (2001c). Prediction in complex communities: analysis of empirically derived Markov models. *Ecology*, 82(2):580–598.

Wootton, J. T. (2004). Markov chain models predict the consequences of experimental extinctions. *Ecology Letters*, 7(8):653–660.

Appendix: optimization methods

Here, we describe the optimization methods used to find maximum likelihood parameter estimates for the LV and linear models.

Parameter transformations and initial guesses

Finding the maximum likelihood estimate $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta})$ is much easier if $\boldsymbol{\theta} \in \mathbb{R}^p$ for a p-dimensional parameter, because we can then use an unconstrained optimization method. The original parameters are constrained. For example $0 < p_i < 1$ and $\sum_i p_i = 1$ for the initial conditions, and $a_{ij} > 0$ for interaction coefficients in the LV model. In the LV model we therefore transform to the unconstrained parameters $\eta_i = \log(p_i(0)/p_s(0))$, $1 \leq i \leq s-1$ (Bickel and Doksum, 2001, p. 55) for the initial conditions, and $\log a_{ij}$ for the interaction coefficients. Optimization is an iterative process requiring initial guesses at parameter values. We set the initial state frequency guesses to $p_i(0) = (n_i(0) + 1)/\sum_i (n_i(0) + 1)$, rather than the obvious $n_i(0)/\sum_i n_i(0)$. This is because if any initial frequencies are zero in the LV model, these states will never appear at subsequent time intervals. We used uniform $(0, 1)$ pseudorandom numbers for initial guesses at a_{ij}.

For the linear model, the initial state probabilities have no effect on the estimates of transition rates. We therefore know that the maximum likelihood estimates for this model are $\hat{p}_i(0) = n_i(0)/\sum_i n_i(0)$, and we can treat them as fixed when estimating the q_{ij}. We used uniform $(0, 1)$ pseudorandom numbers for initial guesses at q_{ij} (as above, we used a log transform to ensure $q_{ij} > 0$).

Implementation

For the linear model, we used the BFGS quasi-Newton optimization algorithm with mixed cubic and quadratic line search implemented as function `fminunc` in the Matlab Optimization Toolbox version 3.1, with Matlab R2006b (The Mathworks, Inc., Natick, MA). This algorithm is not guaranteed to find a global optimum, so we ran the optimization ten times from different random initial guesses, and chose the result with the best likelihood. We also experimented with a genetic algorithm to find good initial guesses for optimization (Matlab Genetic Algorithm and Direct Search Toolbox version 2.0.2), but did not get better results. For the LV model, the initial Matlab implementation was too slow, so we wrote C code to call the NAG FORTRAN library version 21 for Linux (Numerical Algorithms Group, Oxford). We used the quasi-Newton optimizer `E04JYF`, and the stiff ordinary differential equation solver `D02EJF`. Again, we chose the best of ten optimizations from random initial guesses. Optimizations were done on a Linux workstation with an Intel Xeon 3 GHz processor and 1G RAM. Ten replicate optimizations of the linear model took less than 10 minutes for the data analyzed below, while ten replicate optimizations of the LV model took one to four hours. In most cases, convergence of the optimization was not entirely successful. For the linear model, the line search step often failed before

the optimizer had converged, although usually the gradient at the final estimate was quite small. For the LV model, we often encountered numerical problems with solving the differential equations, forcing us to abandon the optimization while the gradient was still fairly large. This was probably because some transition probabilities $\mathbf{p}_j$ (Equation 15) were extremely small. Thus, although we know that we can estimate parameters with reasonable accuracy (see below), we cannot use the inverse of the Fisher information matrix as an estimate of the covariance matrix.

Code for both models is available at `http://www.liv.ac.uk/~matts/`.

Performance

We carried out initial experiments to determine whether to aggregate states. We estimated parameters for the LV model as above, simulated using the best estimates and the number of points present in the first real sample, and re-estimated parameters from the simulated data. The Pearson correlation between true and estimated parameters was 0.98. However, the slope of the least-squares regression between true and estimated parameters (which should be 1) was significantly less than 1 (0.86, 95% confidence interval [0.81, 0.90]). The intercept should be 0, and had a wide confidence interval but was not significantly different from 0 (0.55, 95% confidence interval [-0.24, 1.34]). High coefficients were consistently underestimated, perhaps because the likelihood surface becomes quite flat when some coefficients are very large. Since these coefficients are likely to be of interest, we aggregated all the acroporid corals into a single state, reducing the number of parameters from 80 to 35 and making the optimization problem easier. Repeating the estimation test, we obtained a regression slope that did not differ significantly from 1 (0.96, 95% confidence interval [0.91, 1.00]) and an intercept that did not differ significantly from 0 (0.23, 95% confidence interval [-0.51, 0.96]). Furthermore, the true parameters were within the 95% confidence interval (likelihood ratio for comparison between true and estimated parameters, $\Delta l = 17.88$, $\mathrm{df} = 35$, $p = 0.43$).

We carried out further simulations using data generated from the estimated LV, linear and time-averaged discrete models to check the performance of the estimation. We generated 20 data sets under each model, and estimated parameters for all models for each data set as above (in each case, selecting the best of ten optimization replicates for each simulation replicate, as was done with the real data). Linear optimizations were done using Matlab R2006a on a Sun Fire V880 with eight UltraSPARC III processors. LV optimizations were done on 20 AMD Opteron 2.2 GHz processors in the NW-GRID cluster. Table 5 shows the performance of AIC in selecting the correct model in each case. The totals do not sum to 20 because only data sets for which all models produced an estimate are included. The LV optimization failed completely in one case when linear was the true model and one case when time-averaged discrete was the true model. When LV was the true model, we excluded five data sets for which the best optimization replicate terminated in less than 120 seconds without finding a minimum, compared to an average of 3792 seconds in the other

replicates. Such cases produced very poor results but are easy to detect and were not observed for the real data. The only potential problem with model identification was when the linear model was the true model. In this case, the LV model was selected in 4/19 cases. This is probably because for the linear parameters estimated here, the system quickly approaches an equilibrium (Figure 1). As discussed in 'linear Markov models', if there is an LV model with the same equilibrium, it will be difficult to distinguish from the linear model. This is unlikely to be a problem for the real data, which do not appear to be close to equilibrium (Figure 1).

Table 6 shows the Pearson correlations between true and estimated transition coefficients or probabilities, and the slopes and intercepts of the corresponding regressions. For all models, there is a high correlation between true and estimated parameters. For the LV and time-averaged discrete models, the mean regression slope and intercept were close to one and zero respectively, showing that estimated parameters were close to their true values. However, in the linear model, the mean slope and intercept were very different, because of a few very large rate estimates in some replicates. This is probably due to the potential identifiability problems for the linear model parameters mentioned in the Results. We carried out likelihood ratio tests comparing the true parameters with the maximum likelihood estimates from each set of simulations. When the true model was LV (35 degrees of freedom), the true parameters were not rejected at the 5% level in any of the 15 replicates that completed. When the true model was linear (30 degrees of freedom, considering the rate estimates only, with initial frequencies fixed at the ML values), the true parameters were not rejected in any of the 20 replicates. When the true model was time-averaged discrete (30 degrees of freedom, transition probabilities only), the true parameters were rejected in 1 of 20 replicates.

In summary, we are reasonably confident that we can get good parameter estimates and correct model identification for these data.

Table 1: Log likelihoods (l) and Akaike's Information Criterion (AIC) for Protected Crest models, ordered by increasing AIC.

Model	l	Parameters	AIC
Saturated discrete	-1.5413×10^4	485	3.1797×10^4
LV	-1.6208×10^4	35	3.2486×10^4
Linear	-1.7184×10^4	35	3.4437×10^4
Time-averaged discrete	-1.7330×10^4	35	3.4730×10^4

Table 2: Estimated **A** matrix (years^{-1}) for LV model, Protected Crest. Estimated initial state frequencies were [0.4938, 1.0441 $\times 10^{-5}$, 0.0174, 0.0102, 0.0175]$^{\mathrm{T}}$.

	1	2	3	4	5	6
1: acroporid corals	0	0.6075	6.3557	0.0802	0.1525	0.9572
2: soft corals	0.3654	0	1.13e-54	0.8443	62.1711	0.5078
3: algae	28.9099	2.20e-25	0	11.8025	6.94e-38	25.6192
4: massive corals	0.7293	2.79e-47	0.4187	0	0	0.8297
5: pocilloporid corals	8.40e-12	4.21e-04	38.6881	1.12e-06	0	1.3899
6: free space	0.3068	0.1381	23.1526	0.4922	1.0719	0

Table 3: Estimated **Q** matrix (years^{-1}) for linear model, Protected Crest. Estimated initial state frequencies were [0.4664, 0, 0.0443, 0.0103, 0.0032, 0.4759]$^{\mathrm{T}}$.

	1	2	3	4	5	6
1: acroporid corals	-0.6156	2.42e-06	23.8709	1.40e-07	0.1930	5.96e-04
2: soft corals	8.14e-08	-0.1613	8.52e-04	0.0077	4.34e-07	0.0119
3: algae	0.6130	0.1613	-172.8283	0.7583	0.0013	1.7828
4: massive corals	7.42e-08	1.58e-05	1.4256	-0.7662	2.29e-06	0.0292
5: pocilloporid corals	2.68e-06	6.08e-07	0.0040	2.63e-08	-0.9342	0.0057
6: free space	0.0026	2.79e-06	147.5269	1.37e-04	0.7399	-1.8302

Table 4: Estimated **P** matrix (transition probabilities ignoring variation in time interval) for time-averaged discrete model, Protected Crest. Estimated initial state frequencies were [0.4664, 0, 0.0443, 0.0103, 0.0032, 0.4759]$^{\mathrm{T}}$.

	1	2	3	4	5	6
1: acroporid corals	0.5314	0.0280	0.2322	0.1321	0.1974	0.2025
2: soft corals	0.0038	0.8349	0.0047	0.0111	0	0.0144
3: algae	0.0081	0	0.0332	0.0153	0	0.0068
4: massive corals	0.0133	0.0062	0.0521	0.3408	0	0.0318
5: pocilloporid corals	0.0013	0	0	0	0.2237	0.0038
6: free space	0.4420	0.1308	0.6777	0.5007	0.5789	0.7407

Table 5: Model selection using Akaike's Information Criterion from replicate data sets simulated using the parameter estimates for the Protected Crest site.

True model	Selected model			
	saturated	LV	linear	time-averaged discrete
LV	0	15	0	0
linear	0	4	15	0
time-averaged discrete	0	0	0	19

Table 6: Quality of transition parameter estimation from n replicate data sets simulated using the parameter estimates for the Protected Crest site. Numbers are means, with standard deviations in parentheses.

True model	n	correlation	slope	intercept
LV	15	0.95 (0.06)	1.10 (0.36)	-0.12 (1.41)
linear	20	0.98 (0.03)	1.14×10^3 (4.86×10^3)	-937 (4.00×10^3)
time-averaged discrete	20	0.997 (0.003)	1.01 (0.03)	3.89×10^{-4} (0.002)

Figure 1: Protected Crest data (circles), LV model (solid line), linear model (dashed line), and time-averaged discrete model ignoring variation in time intervals (crosses). States are (a) acroporid corals, (b) soft corals, (c) algae, (d) massive corals, (e) pocilloporid corals, (f) free space. Time is measured in years since the first sample. The y axis scales are different in each panel.