Lotka-Volterra competition models for sessile organisms

Matthew Spencer
School of Biological Sciences, University of Liverpool, Liverpool, UK
m.spencer@liverpool.ac.uk

Jason E. Tanner
SARDI Aquatic Sciences, Henley Beach, Australia and School of Earth and Environmental Sciences, University of Adelaide, Adelaide, Australia
Tanner.Jason@sa.gov.sa.gov.au

Received _______________; accepted ________________

Short title: models for sessile organisms
ABSTRACT

Markov models are widely used to describe the dynamics of communities of sessile organisms, because they are easily fitted to field data and provide a rich set of analytical tools. In typical ecological applications, at any point in time, each point in space is in one of a finite set of states (e.g. species, empty space). The models aim to describe the probabilities of transitions between states. In most Markov models for communities, these transition probabilities are assumed to be independent of state abundances. This assumption is often suspected to be false, and is rarely justified explicitly. Here, we start with simple assumptions about the interactions among sessile organisms, and derive a model in which transition probabilities depend on the abundance of destination states. This model is formulated in continuous time and is equivalent to a Lotka-Volterra competition model. We fit this model and a variety of alternatives in which transition probabilities do not depend on state abundances to a long-term coral reef data set. The Lotka-Volterra model describes the data much better than all models we consider other than a saturated model (a model with a separate parameter for each transition at each time interval, which by definition fits the data perfectly). Our approach provides a basis for further development of stochastic models of sessile communities, and many of the methods we use are relevant to other types of community. We discuss possible extensions to spatially explicit models.

Subject headings: Markov models, coral reefs, Lotka-Volterra competition, maximum likelihood, statistical modelling
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 influential Roughgarden et al. (1985) model of an open population with space-limited recruitment is also a linear Markov model in its original form, although this not immediately obvious.

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 matrix population models (Caswell 2001), many which can equally be applied to Markov models of communities. While Markov models are very simple, in at least some cases they have proven to have surprisingly good predictive ability (Wootton 2004).

In the Markov models considered here, 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 (although Markov models with an infinite set of possible states also exist). Conditional on the current state, future states are independent of the past. 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. Nonlinear open-population models have been developed, in which either mortality (Roughgarden et al. 1985) or growth (Muko et al. 2001a) depend on the amount of free space. There is empirical evidence for such effects in some but not all populations to which they have been applied (Gaines and Roughgarden 1985; Hyder et al. 2001; Muko et al. 2001b; Svensson et al. 2004). 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 preparation) 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.

Most of these density-dependent models are formulated in discrete time (exceptions include Muko et al. 2001a,b). 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 (a saturated model is one with a separate parameter for each transition probability at each time interval, which by definition fits the data perfectly). 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.

The models we evaluate are not fundamentally novel. However, in theoretical ecology, the effort expended on developing new models far exceeds that expended on evaluating how well these models describe real data. We agree with the idea that “the confrontation between more than one model arbitrated by the data underlies science” (Hilborn and Mangel 1997, p. xv). As far as we are aware, this is the first attempt to formally compare the fit of linear Markov models and nonlinear models to a field data set, although comparisons have been made between the results of different linear models (e.g. Tanner et al. 1996).

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
that this is too difficult or because they are of secondary interest).

2. That conditional on the current state, future states are independent of the past (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 rates of transitions are defined by a constant generator matrix (e.g. Norris 1997, section 2.1). This assumption can be tested by fitting models in which transition rates are either dependent on or independent of the proportion of $i$, and using model comparison methods to select the better model, as described in Appendix A4. Our treatment of density dependence seems natural for situations where colony growth is the main mechanism by which transitions to a non-empty state occur. In support of this view, J. E. Tanner et al. (in preparation) show that for the same data set, transition probabilities from $j$ to $i$ in discrete time depend on the proportion of $i$ much more frequently than on the proportion of $j$.

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.
Although J. E. Tanner et al. (in preparation) found that discrete-time transition probabilities to empty space did depend on the proportion of empty space for some species groups, these relationships were relatively weak. 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.

5. That interaction coefficients are constant over time. In particular, we are ignoring seasonal effects. Seasonality may be important in some systems, and can in principle be incorporated into models of the kind we develop. This might be useful because some aspects of coral demography, such as annual mass-spawning (Harrison and Wallace 1990), can be strongly seasonal. However, colonization of free space in this system is overwhelmingly dominated by growth of existing colonies rather than recruitment, and the transition rates do not separate recruitment from colony growth. Mortality due to cyclones is also likely to be seasonal. However, cyclones are rare, and except for extreme catastrophic events, it is difficult to distinguish cyclone effects from routine mortality that occurs throughout the year (Tanner et al. 1994). From a pragmatic perspective, it might be difficult to estimate seasonally varying parameters because our data were collected at intervals greater than one year. Our aim here is to produce a simple model which can be tested using field data, with which we can evaluate the influence of state frequencies on interaction rates. However, seasonal effects may be very important in other sessile assemblages, such as temperate mussel communities (Wootton 2001c).

6. That the system is of infinite spatial extent and that local spatial effects are unimportant. These assumptions allow us to develop a deterministic mean-field differential equation model for the dynamics of the whole system, which governs the
transition rates in a Markov process representing the states of a finite number of sampled points in space. In reality, local spatial effects are likely to be important. In the Conclusions, we consider some ways in which they could be modelled.

We formulate the resulting model at the community level 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

In Appendix A1, we describe a stochastic model for the rate of transitions from state $j$ to state $i$ at a point in space, as a function of the number of points in state $i$. In the absence of detailed information on the spatial arrangement of points and dispersal distances, we then use a mean-field approximation in which the rate of transitions from $j$ to $i$ per unit frequency of state $j$ is $a_{ij}x_i$, where the coefficient $a_{ij}$ has dimensions $T^{-1}$, and $x_i$ is the dimensionless frequency of state $i$. For transitions to empty space, we assume that there is no dependence on the frequency of empty space, and model the rate per unit frequency of $j$ as $a_{ej}$ (dimensions $T^{-1}$). 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:

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

with $x_i \geq 0$ and $\sum_{i=1}^{s} x_i = 1$.

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 $j$. If such terms were present, comparing $a_{ji}$ with $a_{ie,j}$ would tell us about the relative importance of direct negative effects of $j$ on $i$ versus enhancement of colonization by $i$ in the presence of $j$.

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

$$\frac{dx}{dt} = (XA - C)x = R(x)x$$

where $R(x)$ 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 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.

The general Lotka-Volterra competition model is

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

(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 $\alpha_{ik}$ (dimensionless) is an interspecific
competition coefficient measuring the effect of species \( k \) on the growth rate of species \( i \).

Note that as in Eq. 1, the \( x_i \) are proportions.

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 Eq. 1 with \( i \neq e \) and rearranging, we obtain

\[
\frac{dx_i}{dt} = (a_{ie} - a_{ei})x_i - a_{ie}x_i^2 - \sum_{k \neq e, i} (a_{ki} + a_{ie} - a_{ik})x_ix_k
\]

which is identical to Eq. 3 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).

If Eq. 2 is at equilibrium, \( R(x) \) does not vary over time. Let \( x^* \) be an equilibrium state of Eq. 2, and set \( Q = R(x^*) \). Eq. 2 is therefore indistinguishable at equilibrium from the homogeneous continuous-time linear model with transition rates

\[
\frac{dx}{dt} = Qx
\]

where \( Q \) 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. Note that if the system is not at equilibrium, Eq. 2 and Eq. 5 will behave differently. 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

\[
x(T + t) = P(t)x(T)
\]
where \( x(T) \) is a vector of state probabilities at time \( T \), and \( P(t) \) is a transition probability matrix whose \( ij \)th 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 with generator \( Q \), then

\[
P(t) = \sum_{m=0}^{\infty} \frac{(Qt)^m}{m!} = e^{Qt}
\]

(7)

where \( e^{Qt} \) is a matrix exponential. \( P(t) \) 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): the condition for this is that \( P \) is regular (Kijima 1997, p. 52).

In many ecological analyses, the \( P \) 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 \( P \) matrix is independent of state frequencies, or that the frequencies are close to equilibrium.

4. Likelihood, parameter estimation and model selection

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 models. Suppose we have a sequence of states \( y_0, y_1, \ldots, y_k \) 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

\[
P(y_0, y_1, \ldots, y_k) = P(y_k|y_{k-1})P(y_{k-1}|y_{k-2})\ldots P(y_1|y_0)P(y_0)
\]

\[= P(y_0) \prod_{m=1}^{k} P(y_m|y_{m-1})
\]

(8)
where $P(y_m|y_{m-1})$ is the probability of observing state $y_m$ at time $t_m$ given state $y_{m-1}$ at time $t_{m-1}$ and $P(y_0)$ 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

$$L = \prod_{h=1}^{v} P(y_{0,h}) \prod_{m=1}^{k} P(y_{m,h}|y_{m-1,h})$$

$$= \left[ \prod_{j} p_j(0)^{n_j(0)} \right] \prod_{m=1}^{k} \prod_{ij} p_{ij}(m, m-1)^{n_{ij}(m,m-1)}$$

where $y_{m,h}$ 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

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

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.

In deriving Eq. 10, we assumed that the sample points have negligible effects on the dynamics of the entire system. This is reasonable if the spatial extent of the system is infinite. We also assumed that the sample points are independent of each other. This is more problematic, unless sample points are far apart relative to the scale of spatial
interactions (which is unlikely to be strictly true for our data). However, 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. If there are dependencies, it will be as though there were fewer independent observations than sample points, so hypothesis tests should be interpreted cautiously. In the Conclusions, we suggest some possible solutions to the problem of spatial dependencies.

In Appendix A2, we give the transition probabilities for the LV model, the linear homogeneous continuous-time model, a saturated discrete-time model with one parameter for every combination of time interval, source state and destination state, and a time-averaged discrete-time model with the same transition probabilities for every time interval.

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 Appendix A3, we illustrate how small singular values of the Jacobian matrix of partial derivatives of expected values of each category of observations with respect to each parameter can indicate identifiability problems (Catchpole and Morgan 1997). We apply this criterion to all the models we consider.

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 in Appendix A2, Eqs. A.6 or A.7. We do not have closed-form estimates for the other models, so we use numerical optimization as described in Appendix A5.

Appendix A4 describes the methods we used to compare models. Briefly, 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). We can also use Akaike’s Information Criterion (AIC) to compare all four models with each other (Akaike 1992). The Akaike weight for each model can be treated as an estimate of the probability that it is the best in the set under consideration (Burnham and Anderson 2004).

5. 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. In parallel with the current study, a discrete-time density-dependent model is also being developed, although it is not in a format that allows direct comparison to the models presented here. Data were collected from photographs of fixed 1m$^2$ quadrats taken at unequal intervals over 27 years, from 1962 to 1989. Grids of points, referenced to fixed stakes, were placed over the photographs and the species present at each point recorded as described in Tanner et al. (1994). Because the locations of grid points were fixed, it was possible to determine how the species present at points in space changed over time, giving counts of transitions for each time interval.

72 species of corals and 9 species of algae were observed in the quadrats over the 27-year study period. In previous studies, these were grouped into eight categories based on taxonomy and morphology, plus a free space state (Tanner et al. 1994, 1996). Free space was usually occupied by organisms such as crustose coralline and turfing algae, but is available for colonization by corals and macroalgae (J. E. Tanner et al., in preparation). Initial analyses (Appendix A5) showed that the reliability of parameter estimation was improved by aggregating the four original acroporid coral states into a single state, reducing
the number of parameters from 80 to 35. All the models presented here, including the
saturated model, are therefore based on six categories: three hard coral groups (acroporids,
massives, and pocilloporids); soft corals; algae; and free space. 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
A5) 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 Pools). However, these had fewer sample points in time and/or space, and
simulation studies like those described in Appendix A5 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.

6. Results and discussion

6.1. 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. J. E. Tanner et al. (in preparation) also show that discrete-time transition probabilities can be strongly density-dependent in this system, with up to 74% of the variance in some transition probabilities being explained by density.

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 Eq. 2 for the LV model, and Eq. 5 for the linear model. For the time-averaged discrete model we generated expected frequencies using Eq. 6, 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 $1 \times 10^{-9}$, which may indicate potential identifiability problems (Appendix A3). 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 $1 \times 10^{-5}$. We can therefore be more confident that the parameter estimates for the LV model have biological meaning.

### 6.2. Interaction coefficient estimates in the LV model

Parameter estimates for all the non-saturated models are given in Table 2 (LV model) and Appendix: Tables A3 and A4 (linear and time-averaged discrete 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 $1 \times 10^{-9}$ 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 preparation).

We also investigated the long-term dynamics of the LV model. Using the estimated interaction coefficients and initial conditions from Table 2, we solved the system numerically for 10000 years. The state frequencies approached an equilibrium with 46% acroporids, 35% free space, 11% soft corals, 8% massive corals, and no pocilloporid corals or algae. There were transient oscillations, which were negligible after about 1000 years. We analyzed the local stability of this equilibrium by examining the eigenvalues of the Jacobian (e.g. May 1971). These eigenvalues all had negative real parts. This confirms that the equilibrium is locally stable (and that at this equilibrium, we could find a linear model that would be indistinguishable from the LV model).
7. 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. Density-independent colonization is the distinguishing feature of open population models (Roughgarden et al. 1985), and is typically associated with long-distance larval dispersal. It is implausible that the system we studied is completely closed, but it does seem likely that local abundances are important in determining transition rates. This reflects the fact that colonization by recruitment of new individuals is rare compared to colonization by lateral growth of existing individuals.

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 systems 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 increased 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 spatially explicit. 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. Gratzer et al. (2004, p. 9) provide an interesting discussion of the consequences of non-random distributions of individuals for models of spatial interaction in forest systems, phrased in terms similar to those we used in deriving the LV model (Section 2.2). Although there are spatially explicit models of sessile marine communities (e.g. Wilson et al. 1996; Burrows and Hawkins 1998; Wootton 2001b; Robles and Desharnais 2002; Langmead and Sheppard 2004; Dunstan and Johnson 2005, 2006), little effort has yet been expended on fitting to time series, or performance relative
to non-spatial models. Instead, most models have focussed on qualitative patterns. One way to reduce the complexity of these models is to use moment approximations to obtain more accurate differential equation models than the mean-field system we considered (e.g. Bolker and Pacala 1997; Pascual and Levin 1999). Alternatively, spatially explicit models are often formulated as stochastic cellular automata, which are linear Markov models with a very large state space. The likelihood for a discrete-time cellular automaton is easy to calculate, because the states of cells at time $t+1$ are conditionally independent given the states of their neighbourhoods at time $t$ (Balzter et al. 1998; Craiu and Lee 2006). The problem is more difficult in continuous time, because events outside the neighbourhood of a cell may affect its future state through a sequence of changes to intervening cells. The full generator matrix for such a model will usually be too large to store or exponentiate. Possible approaches include Monte Carlo estimation of likelihoods by importance sampling (Juneja and Shahabuddin 2001), pseudolikelihood methods (Besag 1975), and approximate Bayesian computation (Sisson et al. 2007). Perhaps the most closely related problem is that of estimating parameters for a presence-absence metapopulation from time series data. If the metapopulation is not observed every year, Bayesian data augmentation methods can be used to integrate over sample paths for unobserved intervals (O’Hara et al. 2002). However, the problem is more difficult in continuous time because the length of the sample path is unknown. An alternative approach is to treat a single observed spatial pattern as a sample from the stationary distribution of a continuous-time Markov chain, as has been done for models of forest gap creation and regeneration (Schlicht and Iwasa 2004). It is not clear that this would be appropriate for our system, given the large variation in state frequencies over time.

In conclusion, the models we have investigated here are simple, and have already been well-studied in theoretical ecology. Our main contribution is the attempt to compare the performance of these models as quantitative descriptions of long-term field data.
Stochastic models of real time series can both increase our understanding of the underlying mechanisms, and help us to make better predictions about the future dynamics of populations and communities (e.g. Foley 1994; Kendall et al. 1999; Thomas et al. 2005).

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. The manuscript was improved by detailed and perceptive comments from three anonymous reviewers. We are also grateful to Steve Brooks and Damian Clancy for helpful suggestions, Don DeAngelis for comments on an earlier draft, and Mary Thorp and Cliff Addison for technical support.
REFERENCES


survival in populations of *Semibalanus balanoides*. Marine Ecology Progress Series, 275:185–197.


Table 1: Log likelihoods ($l$), number of parameters and Akaike’s Information Criterion (AIC) for Protected Crest models, ordered by increasing AIC.

<table>
<thead>
<tr>
<th>Model</th>
<th>$l$</th>
<th>Parameters</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturated discrete</td>
<td>$-1.5413 \times 10^4$</td>
<td>485</td>
<td>$3.1797 \times 10^4$</td>
</tr>
<tr>
<td>LV</td>
<td>$-1.6208 \times 10^4$</td>
<td>35</td>
<td>$3.2486 \times 10^4$</td>
</tr>
<tr>
<td>Linear</td>
<td>$-1.7184 \times 10^4$</td>
<td>35</td>
<td>$3.4437 \times 10^4$</td>
</tr>
<tr>
<td>Time-averaged discrete</td>
<td>$-1.7330 \times 10^4$</td>
<td>35</td>
<td>$3.4730 \times 10^4$</td>
</tr>
</tbody>
</table>
Table 2: Estimated $A$ matrix (years$^{-1}$) for LV model, Protected Crest.

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

Note: Estimated initial state frequencies were $[0.4938, 1.0441 \times 10^{-5}, 0.0174, 0.0102, 0.0175, 0.4612]^T$. 
Fig. 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 vertical axis scales are different in each panel.