Lotka-Volterra competition models for sessile organisms: appendix

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@saugov.sa.gov.au

Received _____; accepted _____

A1. Derivation of the LV model

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

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

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

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

If there are spatial effects, we can justify Eq. A.2 as the first term in a Taylor series approximation. Our approximation is very similar to the mean-field version of an interacting particle system (Durrett and Levin 1998). The technical conditions under which we can make the connection between a stochastic, mechanistic model and a system of deterministic ordinary differential equations at the community level are given in Kurtz (1970, 1971).

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 show in Appendix A2.1, 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

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

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.

A2. Transition probabilities

A2.1. Transition probabilities in the LV model

We now need to calculate $p_{ij}(m, m-1)$ for the model specified by Equation 1. 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 1):

$$\frac{dp_{ij}}{dt} = \begin{cases} -\left(a_{ei} + \sum_{k \neq e,i} a_{ki} x_k\right) p_{ij} + \sum_{k \neq i} a_{ik} x_i p_{kj} & i \neq e \\ -\sum_{k \neq e} a_{ke} x_k p_{ej} + \sum_{k \neq e} a_{ek} p_{kj}, & i = e \end{cases} \tag{A.4}$$

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 A.4 in matrix form by substituting $\mathbf{p}_j$ for $\mathbf{x}$ (the unconditional probabilities of each state) in Equation 2:

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

We can integrate Equations 2 and A.5 numerically to get the transition probabilities for any time interval, and insert these into Equation 10 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.

A2.2. Transition probabilities in the linear model

For a homogeneous continuous-time linear Markov model (Equation 5), 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).

A2.3. 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

$$\hat{p}_{ij}(m, m-1) = \frac{n_{ij}(m, m-1)}{\sum_k n_{kj}(m, m-1)} \tag{A.6}$$

(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.

A2.4. 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

$$\hat{p}_{ij} = \frac{\sum_m n_{ij}(m, m-1)}{\sum_m \sum_k n_{kj}(m, m-1)} \tag{A.7}$$

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

A3. Identifiability

An identifiable model is one for which $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \implies l(\boldsymbol{\theta}) \neq l(\boldsymbol{\theta}_0)$, 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

$$\mathbf{D} = \begin{bmatrix} \partial y_1/\partial\theta_0 & \partial y_2/\partial\theta_0 & \ldots & \partial y_n/\partial\theta_0 \\ \partial y_1/\partial\theta_1 & \partial y_2/\partial\theta_1 & \ldots & \partial y_n/\partial\theta_1 \\ \partial y_1/\partial\theta_2 & \partial y_2/\partial\theta_2 & \ldots & \partial y_n/\partial\theta_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ x_1 & x_2 & \ldots & x_n \end{bmatrix} \tag{A.8}$$

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.

A4. Model comparison

For comparing the saturated with a simpler model, the test statistic $2(l_{\text{saturated}} - l)$ has an asymptotic $\chi^2_{\Delta p}$ distribution if the simpler model is correct, where $l_{\text{saturated}}$ is the log likelihood of the saturated model, l is the log likelihood of the simpler model, and Δp is the difference in the number of parameters between the saturated model and the simpler 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 $l_k = \exp((\mathrm{AIC}_0 - \mathrm{AIC}_k)/2)$, 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).

A5. Optimization methods

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

A5.1. 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 $q_{ij} > 0$).

A5.2. 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 A.5) 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/`.

A5.3. 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, $\Delta l = 17.88$, 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 A1 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 A2 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 from one and zero, 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.

REFERENCES

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

Bickel, P. J. and K. A. Doksum. 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**:345–370.

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

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

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

Durrett, R. and S. Levin. 1998. Spatial aspects of interspecific competition. Theoretical Population Biology, **53**:30–43.

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

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

Kurtz, T. G. 1970. Solutions of ordinary differential equations as limits of pure jump Markov processes. Journal of Applied Probability, **7**:49–58.

225 Kurtz, T. G. 1971. Limit theorems for sequences of jump Markov processes approximating
226 ordinary differential equations. Journal of Applied Probability, **8**:344–356.

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

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

Table A1: 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 A2: Quality of transition parameter estimation from n replicate data sets simulated using the parameter estimates for the Protected Crest site.

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)

Note: numbers are means, with standard deviations in parentheses. Correlation, slope and intercept are for the relationship between true and estimated parameters.

Table A3: Estimated **Q** matrix (years^{-1}) for linear model, Protected Crest.

	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

Note: Estimated initial state frequencies were $[0.4664, 0, 0.0443, 0.0103, 0.0032, 0.4759]^{\mathrm{T}}$.

Table A4: Estimated **P** matrix (transition probabilities ignoring variation in time interval) for time-averaged discrete model, Protected Crest.

	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

Note: Estimated initial state frequencies were $[0.4664, 0, 0.0443, 0.0103, 0.0032, 0.4759]^{\mathrm{T}}$.