

Jeremy W. Fox, William A. Nelson, and Edward McCauley. 2010.
Coexistence mechanisms and the paradox of the plankton: quantifying
selection from noisy data. *Ecology* 91:1774–1786.

APPENDIX A. Derivation and estimation of selection metrics.

Derivation of the general coefficient of selection equation (main text Eq. 1)

Consider a population of n competitors. These competitors could be distinct species, or asexually reproducing genotypes; the derivation holds in either case. The coefficient of selection $s_i(t)$ for competitor type i ($i = 1, \dots, n$) can be defined as the difference between its per-capita growth rate $r_i(t)$ and that of the rest of the population $\hat{r}_i(t)$:

$$s_i(t) = r_i(t) - \hat{r}_i(t). \quad (\text{A.1})$$

Starting with the first term in Eq. A.1, the per capita growth rate of competitor i is given as

$$r_i = \frac{1}{N_i} \frac{dN_i}{dt}, \quad (\text{A.2})$$

where N_i is the abundance of competitor i in the population (time index dropped for clarity). Using the fact that $N_i = p_i N$, where N is the total abundance of the population and p_i is the relative abundance of competitor i in the population, the per-capita growth rate (Eq. A.2) can be expressed as

$$\begin{aligned}
r_i &= \left(p_i \frac{dN}{dt} + N \frac{dp_i}{dt} \right) \frac{1}{N_i} \rightarrow \\
r_i &= \left(p_i \frac{1}{p_i N} \frac{dN}{dt} + \frac{N}{N_i} \frac{dp_i}{dt} \right) \rightarrow, \\
r_i &= r + \frac{dp_i}{dt} \frac{1}{p_i}
\end{aligned} \tag{A.3}$$

where r is the per-capita growth rate of the entire population.

The second term in Eq. A.1 is the per-capita growth rate for the sub-population that does not include competitor i . Starting with the definition of the sub-population

$\hat{N}_i = N - N_i$, the second term can be expressed as

$$\hat{r}_i = \frac{1}{\hat{N}_i} \frac{d\hat{N}_i}{dt} = \frac{1}{\hat{N}_i} \left(\frac{dN}{dt} - \frac{dN_i}{dt} \right). \tag{A.4}$$

Using the relation that $\hat{N}_i = N(1 - p_i)$ and Eq. A.3, Eq. A.4 can be expressed as

$$\hat{r}_i = \frac{1}{1 - p_i} (r - p_i r_i) = r - \frac{dp_i}{dt} \frac{1}{1 - p_i}. \tag{A.5}$$

Substituting Eqs. A.3 and A.5 into Eq. A.1, after some algebra yields

$$s_i(t) = \frac{dp_i(t)}{dt} \frac{1}{p_i(t)(1 - p_i(t))}, \tag{A.6}$$

which states that the coefficient of selection is given by the rate of change in the relative abundance of competitor i , adjusted by how close it is to being excluded from the population, or to excluding all others.

Estimating selection dynamics from noisy data

Eq. 1 in the main text is a continuous time model, but observations of proportions are collected at discrete intervals. To contend with this, the traditional approach in evolutionary biology is to use a discrete-time version of Eq. 1 (e.g., Lynch 1987) and directly calculate selection coefficients from the observed change in frequencies. While this yields an estimate of average selection over a time-series (net selection), a priori assumptions about the sampling error distribution and the amount of covariance between successive selection estimates must be imposed to estimate the variation in selection (fluctuating selection). The result is that traditional methods produce biased estimates of fluctuating selection when sampling error is greater than predicted by a Bernoulli distribution (known as overdispersion) or if temporal estimates of selection are autocorrelated. Overdispersion often is observed in count data (White and Bennetts 1996) and many fluctuation-dependent coexistence mechanisms generate autocorrelated selection coefficients.

Here we develop a different approach from traditional methods that relies on fewer assumptions. We begin by fitting a statistical time-series model to competitor proportions that places the error distribution directly on the observed competitor counts as opposed to the derived selection coefficients. Fitting a time-series model also enables us to account for autocorrelation in the selection coefficients and remove the dependence on sampling intervals. To allow for non-stationarity in the time-series, competitor dynamics are

described using a nonparametric regression model that makes no a priori assumptions about temporal dynamics. The statistical model, which assumes a Dirichlet-multinomial distribution ($\text{DM}(\bullet)$) to allow for possibly-overdispersed sampling error (McCullugh and Nelder 1989), is given by

$$n_i(t) \sim \text{DM}(f_i(t), \dots, f_{m-1}(t), N(t)), \quad (\text{A.7})$$

where $n_i(t)$ is the observed count for competitor i ($i = 0, \dots, m$) in a sample of size $N(t)$ taken at time t , and $f_i(t)$ is a cubic spline representing the temporal dynamics. Each competitor has a separate spline function to allow for differential responses, under the constraint that all splines sum to one since we are modeling the dynamics of relative abundance. Since Eq. A.7 is a flexible non-parametric model, model fitting involved estimating the optimal compromise between overfitting (“fitting the noise”) and underfitting using generalized cross-validation (Yee and Wild 1996; Wood 2001; Nelson et al. 2004). Since there are usually more than two competitors in each data set, the statistical problem is multivariate and requires the use of Vectorized General Additive Models (Yee and Wild 1996). The fits were done with the statistical library VGAM (Yee and Wild 1996) in the R software environment (R Development Core Team 2004).

Statistical inference concerning S_f and S_n (defined by Eqs. 3–5 in the main text) is based on parametric bootstrap confidence intervals (Dennis and Taper 1994). These are generated by drawing random samples from Dirichlet-multinomial distributions with the variance structure and parameters estimated from the fitted model (Eq. A.7). Bootstrap samples are drawn with the same sampling frequency and sampling effort as the original

data, and refit in the same manner as the original data. Confidence intervals for net selection and fluctuating selection on each species were estimated from the upper and lower 2.5% percentiles of the bootstrap distribution obtained from 1000 replicates. To test whether fluctuating selection was statistically greater than zero, we multiplied the positive value from Eq. 6b by the sign of net selection for each bootstrap replicate.

References

- Dennis, B., and M. Taper. 1994. Density-dependence in time-series observations of natural populations- estimation and testing. *Ecological Monographs* 64:205–224.
- R Development Core Team. 2004. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: <http://www.R-project.org>.
- Lynch, M. 1987. The consequences of fluctuating selection for isozyme polymorphisms in *Daphnia*. *Genetics* 115:657–669.
- McCullagh, P., and J. A. Nelder. 1989. Generalized linear models, Second Edition. Chapman and Hall, New York, New York, USA.
- Nelson, W. A., E. McCauley, and J. Wimbert. 2004. Capturing dynamics with the correct rates: inverse problems using semiparametric approaches. *Ecology* 85:889–903.
- White G. C., and R. E. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology* 77:2549–2557.
- Wood, S. N. 2001. Partially specified ecological models. *Ecological Monographs* 71:1–25.
- Yee, T. W., and C. J. Wild. 1996. Vector generalized additive models. *Journal of the Royal Statistical Society B* 58:481–493.