Reports

Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models

STEPHANIE E. HAMPTON,[1,4] ELIZABETH E. HOLMES,[2] LINDSAY P. SCHEEF,[1] MARK D. SCHEUERELL,[2] STEPHEN L. KATZ,[3] DANIEL E. PENDLETON,[2] AND ERIC J. WARD[2]

[1]*National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, California 93101 USA*
[2]*Northwest Fisheries Science Center, NOAA Fisheries Service, Seattle, Washington 98112 USA*
[3]*Channel Islands National Marine Sanctuary, NOAA ONMS, Santa Barbara, California 93106 USA*

Abstract. Long-term ecological data sets present opportunities for identifying drivers of community dynamics and quantifying their effects through time series analysis. Multivariate autoregressive (MAR) models are well known in many other disciplines, such as econometrics, but widespread adoption of MAR methods in ecology and natural resource management has been much slower despite some widely cited ecological examples. Here we review previous ecological applications of MAR models and highlight their ability to identify abiotic and biotic drivers of population dynamics, as well as community-level stability metrics, from long-term empirical observations. Thus far, MAR models have been used mainly with data from freshwater plankton communities; we examine the obstacles that may be hindering adoption in other systems and suggest practical modifications that will improve MAR models for broader application. Many of these modifications are already well known in other fields in which MAR models are common, although they are frequently described under different names. In an effort to make MAR models more accessible to ecologists, we include a worked example using recently developed R packages (MAR1 and MARSS), freely available and open-access software.

Key words: food webs; interaction strength; multispecies time series; reactivity; resilience; return rate; state-space models; VAR; vector autoregressive models.

INTRODUCTION

Who eats whom? Which predators have the strongest cascading effects in an ecosystem? Which basal taxa provide the most important resources in a community? What might be the network of ecological repercussions when a particular nutrient decreases, a species is depleted or removed, or when temperature increases? How do these dynamics relate to ecosystem stability?

Diverse approaches to ecosystem modeling have been undertaken to address questions like these, in order to gain better understanding of trophic interactions and potential responses to human perturbations (Plagányi 2007). Such approaches include highly detailed and spatially explicit models, (e.g., ATLANTIS; Fulton et al. 2004), energy flow models based on diet estimates (e.g., EcoSim; Christensen and Walters 2004) or ecophysio-

logical parameters (e.g., Ecosys; Grant 2001), models of ecological communities as networks (Thébault and Fontaine 2010), and individual-based models (e.g., OSMOSE; Shin and Cury 2004). These approaches attempt to develop a mechanistic model of the system, and they require data (or assumptions) about diet, growth and biomass as well as the functional form of the interactions among species. The development of these mechanistic models can be challenging when such data are unavailable and possibilities for cross-validation are limited.

Alternatively, theoreticians, statisticians, and ecologists together have developed and applied statistical models of community dynamics using long-term abundance and environmental data. This approach is grounded in theory concerning the patterns of temporal correlation that emerge from species interactions (Ives 1995, Ives et al. 1999, 2003). Observational time series data are assumed to have been generated from a multivariate autoregressive (MAR) process. The core pieces of this model include a species interaction matrix

[4] E-mail:hampton@nceas.ucsb.edu

quantifying each species' interaction with itself (e.g., density dependence) and interactions with others (e.g., competition, predation), and another covariate matrix quantifying the impact of environmental drivers on species' per capita growth rates. These matrices are estimated from the time series data as an extension of simple linear regression, and the resulting matrices can be converted into estimates of interaction strengths and perturbation responses (Ives et al. 2003). Alternative MAR model structures, representing different functional groupings or presence/absence of specific interactions, can be compared using information theoretic approaches (such as Akaike's information criterion, AIC) that balance model "fit" with complexity (Burnham and Anderson 2010), and comparative metrics of community stability can be computed from the estimated model (Ives et al. 2003, Neubert et al. 2009). The MAR approach facilitates identification of key players and the pathways through which abiotic and biotic drivers affect population dynamics, and also addresses an important complication for other statistical analyses: reciprocal effects (e.g., Hampton and Schindler 2006). Reciprocal effects occur, for example, when predators eat prey "today" and directly reduce prey biomass, but then increase their own biomass "tomorrow" in response to the prey. In this case, the correlation between predator and prey abundances is both positive and negative, but statistical analyses that do not explicitly identify interaction pathways would fail to identify the reciprocal dynamics.

Ecological applications of MAR models are now well documented, and they have been used extensively to understand the dynamics of freshwater plankton communities. However, they have seen limited application outside the study of plankton communities and most ecologists are unlikely to be familiar with these models and their potential. Here we review the existing ecological applications, which are mainly, though not exclusively, in freshwater systems, and then discuss existing and potential extensions of MAR models that may make them more generally applicable for studying ecosystem and community dynamics.

The MAR model formulation

To characterize population growth, a discrete-time Gompertz model is used with log-linear density-dependence (Ives et al. 2003, Dennis et al. 2006). The stochastic univariate Gompertz equation written in log-space is

$$x_{i,t} = bx_{i,t-1} + a_i + w_{i,t} \qquad (1)$$

where $x_{i,t}$ is the natural log density of species i at time step t, a is the intrinsic rate of increase for species i, and b represents the strength of density dependence ($b = 1$ implies density independence, $0 < b < 1$ implies undercompensation, $b < 0$ implies overcompensation). The process error deviation, $w_{i,t}$, represents random deviations resulting from environmental or demographic

stochasticity from one time step to the next. Eq. 1 is a univariate, first-order, autoregressive, or AR(1), process. The AR(1) process can be expanded to include the effects of exogenous variables (i.e., covariates), such as abiotic drivers, on population dynamics. The following equation shows the form for a single time-varying covariate, u, with an interaction strength of c:

$$x_{i,t} = bx_{i,t-1} + a_i + cu_{t-1} + w_{i,t}. \qquad (2)$$

If u were temperature, then c would represent how strongly (positively or negatively) temperature affects log-abundance.

To describe the stochastic dynamics of p interacting species and q environmental covariates, we use a multivariate version of Eq. 2, a MAR(1) model, which serves as a linear approximation for nonlinear stochastic multispecies processes (Ives et al. 2003). Hereafter, we drop the "(1)" for simplicity when referring to the model, and we use "species" to refer to the biotic groupings chosen by the user (e.g., "species" may be a guild of similar species). The MAR model is a system of p linear equations describing the abundances for each species in the community. In matrix form, the MAR model is written as follows:

$$\mathbf{x}_t = \mathbf{B}\mathbf{x}_{t-1} + \mathbf{a} + \mathbf{C}\mathbf{u}_{t-1} + \mathbf{w}_t \qquad (3)$$

where $\mathbf{x}_t$ is the $p \times 1$ vector of log abundances for each of the p species at time t, $\mathbf{a}$ is the $p \times 1$ vector of a values for each species, $\mathbf{B}$ is a $p \times p$ interaction matrix whose elements b_{ij} describe the effect of the density of species j on the per capita growth rate of species i, $\mathbf{u}_{t-1}$ is the $q \times 1$ vector of covariate values at time $t - 1$, and $\mathbf{C}$ is the $p \times q$ matrix whose elements c_{ij} describe the effect of covariate j on species i. The vector of process errors $\mathbf{w}_t$ is assumed to be drawn from a multivariate normal distribution with a mean of 0 and covariance matrix $\mathbf{S}$; the selection of this distribution to describe the random stochasticity that populations experience has its roots in the statistical distribution that emerges from stochastic population dynamics (Ives et al. 2003, Holmes et al. 2007). It is important to note that inclusion of appropriate environmental drivers in $\mathbf{u}$ may be necessary to remove non-stationarity from the residuals (e.g., seasonal patterns). For example, a covariate like temperature, measured properly, might be adequate to remove hourly patterns of activity in a reptile. In other cases, it might be necessary to include fixed seasonal effects using a covariate to specify month or season. Covariates need not be restricted to abiotic drivers. They may be biotic variables that are expected to affect x without exhibiting reciprocal effects in the analysis. For example, we would not expect fish abundance to respond to plankton abundance on a comparable time scale, so an estimated level of fish predation has been incorporated as a covariate in MAR models focused on plankton (Ives et al. 1999, Beisner et al. 2003).

Reports (printed vertically in left margin)

Parameter estimation, parameter uncertainty
and model support

One of the major goals of MAR modeling in ecology has been to estimate interspecific interaction strengths (**B**) in food webs and the effects of abiotic drivers on population dynamics (**C**) by fitting the MAR model to time series of species abundances. The majority of MAR applications have used one of two methods to estimate **B** and **C**: conditional least squares, which requires that data sets have no missing values, and maximum-likelihood. Bayesian approaches have been less frequently used in MAR modeling (cf. Mutshinda et al. 2009), though they can have some advantages. They more easily allow for non-Gaussian error structures or nonlinear process models, but depending on the circumstances, maximum likelihood may perform similarly to Bayesian approaches. Because MAR models are linear models with Gaussian error, the Bayesian solutions are nearly identical to those achieved with maximum likelihood, and Bayesian estimation may take longer to converge.

A second major goal of MAR modeling in ecology has been to reduce the food web model to a set of consistently strong interactions (represented by positive and negative values in **B** and **C**) and a set of least relevant interactions (represented by zeroes in the **B** and **C** matrices), accomplished by a model selection step. While it is possible to use a model with all **B** and **C** elements estimated, these models are typically prohibitively large and the uncertainty in the estimated elements would be high (Ives et al. 1999, 2003, Hampton et al. 2006). In some instances, a priori knowledge of the ecology of the system will help to identify zeroes (e.g., the direct effect of a carnivore on plants), and these zeroes can be set by the user. Otherwise, identification is accomplished with automated randomized search routines and information theoretic metrics, such as AIC (Burnham and Anderson 2010). These search routines construct thousands of MAR models with different **B** and **C** matrices by randomly setting the elements of these matrices to zero, and the models with the lowest AIC scores are identified. After this model search step, additional interactions may set to zero if they do not appear consistently in the candidate models, as defined by some threshold (e.g., 15% of the models [Ives et al. 1999]), or if bootstrapped confidence intervals overlap with zero (Hampton et al. 2006). Several top models may be considered separately, or models can be averaged to arrive at a final model.

The **B** and **C** matrices in the final MAR model can be viewed as a simplified model of the food web structure (species–species interactions) and its environmental drivers (species–environment relationships), respectively. These interaction strengths are driven by the correlation structure in the data and thus their interpretation is subject to the usual constraints on the interpretation of statistical correlations. Some identified relationships may actually be mutual correlations with an unmeasured environmental driver that was not included in the model (Hampton et al. 2006), and some "real" interactions between taxa and their drivers may fail to emerge when they are relatively weak, complex, or inconsistent. Interpretation can be affected by choices about grouping taxa into size classes, trophic guilds or other functional groups, as is common in plankton studies. For example, the tiny freshwater zooplankton *Ascomorpha ovalis* has highly specialized feeding behaviors that allow it to exploit large dinoflagellates (Stelzer 1998), a group of algae mostly inedible to other zooplankton. The otherwise reasonable lumping of *A. ovalis* into "small grazers" and large dinoflagellates into "inedible algae" likely would obscure this relationship, unless these taxa are numerically dominant, in which case one might reach the generally wrong conclusion that small grazers thrive on large dinoflagellates. Similarly, incorrect relationships may emerge in MAR results when particular taxa act as proxies for unmeasured or obscured variables. In cases like these, the structure and interpretation of the model will rely crucially on the natural history expertise of the investigators.

Ecosystem stability metrics from MAR models

In the ecological applications to date, MAR models have been used primarily to assess interaction strengths and infer food web structure, but one can also use MAR modeling to compute stability properties of the system as a whole. Ives et al. (2003) describe a variety of different stability metrics, and we highlight three of these here.

1) The variance of the *stationary distribution* of community states relative to the variance of the process error. This property estimates the variability of the system at equilibrium relative to the process error variability that drives the stochasticity and is determined by the **B** (species interactions) and the **S** (process variances) matrices together. Stability increases as the ratio of the stationary distribution variance to process variance decreases.

2) The *return rate* of the system represents how quickly the community returns to its stochastic equilibrium after a disturbance. The return rate is determined by the dominant eigenvalue of the Kronecker product **B**⊗**B**. Stability increases as return rate increases.

3) The *reactivity* of the system is an index of a community's short-term response to perturbation. The "worst case" reactivity depends only on the eigenvalues of the estimated species interaction matrix **B**. Stability increases as reactivity decreases.

Even for relatively simple ecological systems, interpreting how individual elements of **B** translate into stability may be confusing; however, each of the above metrics offers the advantage of reducing stability to a single number (with confidence intervals). A further advantage is that these metrics are not obscured by the

magnitude of fluctuations. For example, in comparing the dynamics of a highly variable community and a community with low variability, one might immediately conclude that the highly variable community was less stable. Because the highly variable community might be simply reflecting a more variable environmental driver (e.g., temperature), the highly variable community in reality might be just as stable as the community with low variation. When relevant environmental drivers are included in the model, MAR analyses can cut through this overt variability to characterize the fundamental stability of the system; e.g., a MAR may reveal that a community exhibiting more variability is in fact more robust to perturbation than the community exhibiting lower variability.

Previous ecological applications of MAR modeling

In ecology, MAR modeling has been used to infer community dynamics primarily within aquatic sciences. Freshwater plankton communities have been a focus, in part due to frequent data collection and the nature of processing methods (Scheef et al. 2012). Perhaps more importantly, the information content of these time series typically is high; because plankton communities are short-lived, many generations can be represented in a relatively short study.

Previous MAR modeling of plankton data has demonstrated the range of ecological topics and problems that can be addressed using this approach. Analysis of long-term freshwater plankton data using MAR modeling has illuminated direct and indirect pathways through which plankton community dynamics are driven by keystone species. For example, Ives et al. (1999) constructed MAR models of plankton dynamics in a lake where planktivory rates had been manipulated by adding and removing fish, and found that *Daphnia pulex*, an important grazer, strongly influenced other species in the community through both direct and indirect interactions. Hampton and Schindler (2006) discovered an important influence of particular algal taxa (cryptomonads and picoplankton) on zooplankton growth that previously had not been considered when they applied MAR analysis to long-term data from Lake Washington. Several studies have characterized tight competitive or predatory linkages within freshwater plankton communities through MAR analysis. MAR analyses revealed strong predatory effects of an exotic fish on the structures of plankton communities in two Wisconsin lakes (Beisner et al. 2003) as well as strong grazing effects of zooplankton on primary producers in Lake Constance (Huber and Gaedke 2006) and lake mesocosms (Duffy 2007). Use of MAR modeling has facilitated the detection of competitive release between plankton groups during pH manipulations in meso-cosms (Klug et al. 2000) and between *Daphnia* species during fungal parasite epidemics in small kettle lakes (Hall et al. 2009). MAR modeling has been used to characterize the pathways through which various

environmental factors, such as increasing temperature (Hampton et al. 2008), dissolved organic carbon and nutrient levels (Klug and Cottingham 2001), shifting pH (Fischer et al. 2001), and terrestrial carbon input (Carpenter et al. 2005) affect freshwater plankton communities. In these studies, MAR modeling both reinforced existing hypotheses about food web structure and suggested new, unexpected relationships among interacting biota.

In marine systems, there have been several applications of MAR modeling in fisheries sciences. These applications used the terminology vector autoregressive model (VAR), which is the name for a MAR model without covariates (Eq. 3 minus the **c** and **u**) in the econometrics literature. To the best of our knowledge, the first uses of MAR modeling to analyze marine fisheries data were applications involving the forecast of anchovy and sardine populations in the Mediterranean Sea (Stergiou 1991, Stergiou and Christou 1996). Since then, MAR modeling has been used to analyze the effects of climate and catch prices on fishing effort for albacore tuna, Chinook salmon, sablefish, and squid in California's Monterey Bay (Dalton 2001); to investigate the ecosystem linkages that affect the carite, croaker, and honey shrimp fisheries in Trinidad's Gulf of Paria (Dhoray and Teelucksingh 2007); to examine fish declines in the San Francisco Bay estuary (Mac Nally et al. 2010); to assess the portfolio effect in coral fishes (Thibaut et al. 2012); and to evaluate the effects of fishing pressure and environmental fluctuations on dynamics of commercially important cod (*Clupea harengus*) in the Black Sea with additional explorations of alternative management scenarios (Lindegren et al. 2009).

In terrestrial systems, MAR modeling has had limited application. MAR modeling has been used to study lynx-hare dynamics (Vik et al. 2008), to examine the effects of climate change on insect community dynamics (Yamamura et al. 2006), and to compare the relative importance of environmental stochasticity vs. stochasticity driven by interspecific-intraspecific interactions in rodents (Mutshinda et al. 2009). Given that trophic interactions and ecosystem stability are of similar interest across aquatic and terrestrial ecology, and that long-term data sets are available across systems, the scarcity of MAR use outside of freshwater ecology may be due to a lack of familiarity with MAR models and the ways in which they may be modified when data or questions do not conform with existing examples in the ecological literature.

As an illustration, consider the issues encountered when attempting to make the relatively small leap from applying MAR modeling with freshwater plankton systems to doing so with marine plankton systems (Francis et al. 2012, Scheef et al. 2012; Scheef et al., *in press*). In common marine data sets, compared to freshwater data, the investigator likely will find that observation error is higher due to both spatiotemporal

variation in sampling and the more "open" nature of marine systems. There also may be fewer data points available for each taxon, with possibly longer and uneven intervals between collections. These characteristics of multivariate time series are probably representative of a large number of ecological research programs; they do not exclude the possibility of applying MAR modeling, but do necessitate some modifications.

Future directions to broaden MAR usability and applications

Successfully applying MAR models to a broader diversity of long-term ecological data will require extending the MAR framework to explicitly accommodate (1) observation error, (2) spatially replicated time series across large geographic regions, (3) temporally nonuniform sampling, and (4) reduced autocorrelation structure, to accommodate varying "openness" of the system and monitoring schemes with long time lags between samples. State-space MAR models (MARSS) offer a framework for dealing with these issues.

MAR models that do not explicitly account for observation errors (Eq. 3) are limited to data sets where observation error can be safely ignored (Holmes et al. 2012b). When it cannot be ignored, separating process error from observation (measurement) error is critical because ignoring measurement error leads to the well-known consequence of spurious density dependence (Shenk et al. 1998, Knape and de Valpine 2012). In a MAR analysis, this phenomenon manifests as smaller estimates on the **B** diagonal when one uses a MAR analysis without observation error vs. one that includes observation error in the abundance estimates (see the Supplement for an example). Similarly, the off-diagonal terms of the **B** matrix (the inter-species interaction terms) tend to be smaller when observation error is important but not included. State-space MAR models combine a MAR model with an observation error model:

$$y_t = \mathbf{Z}\mathbf{x}_t + \mathbf{D}\mathbf{d}_t + \mathbf{v}_t \qquad (4)$$

where $\mathbf{y}_t$ are the observations and are modeled as a linear function of the hidden species abundances, $\mathbf{x}_t$, and optionally environmental factors that affect observation error, $\mathbf{d}_t$. State-space MAR models are widely used in other fields (e.g., Durbin and Koopman 2012), and have been successfully employed in fisheries research (e.g., Lindegren et al. 2009, Hinrichsen and Holmes 2009). Although state-space MAR models allow one to include an observation model and estimate its variance, estimation of the **B** matrix in the face of unknown observation error is especially challenging (a fact alluded to in Ives et al. 2003) and comes at a cost in terms of higher variance in the **B** estimates. When measurement error variance is small relative to process error variance, a "standard" MAR model with no observation error component may actually improve the overall **B** estimation.

Where observation error is non-negligible, state-space models can be tailored to allow the user to maximally take advantage of the data at hand. For example, spatially distributed observations (spatial "replicates") may be used to facilitate the separation of variances, an approach that has been successful in single-species process models (Humbert et al. 2009). Conceptually, this same approach could be applied to multiple species, treating communities as hierarchical or spatial replicates in a state-space MAR model. If the focus of inference is on species interactions, the process model (Eq. 3) could be constrained to share the **B** matrix across communities, with other parameters optionally shared depending on model assumptions.

State-space MAR models also allow greater flexibility in addressing temporal aspects of ecosystem dynamics—accommodating longer time lags in the process model as well as time-varying parameters. Eq. 3 shows a lag-1 MAR model, where $\mathbf{X}_t$ depends on $\mathbf{X}_{t-1}$. A kth order MAR model, where $\mathbf{X}_t$ depends on $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots \mathbf{X}_{t-k}$, can be re-written as a first-order model by re-defining the **B** matrix and **X** vector (see section 11.3.2 in Tsay 2010). State-space MAR models that include time-varying parameters are known more commonly as dynamic linear models (Shumway and Stoffer 2006). The ability to incorporate time-varying interactions in MAR models may be especially interesting to ecologists, recognizing that species interactions may change with species density, indirect effects, priority effects, a changing environment (Kordas and Dudgeon 2011), or through evolutionary change that occurs on ecologically relevant time scales (Hairston et al. 2005).

Using MAR modeling

Just as linear regression remains valuable to ecologists, even though more complex analytical tools are increasingly available, a standard MAR model will be useful in a variety of cases where the MARSS configuration is unnecessary or inappropriate. Two R packages have been developed that lower the technical barriers to fitting MAR and MARSS models to ecological data. The MAR1 R package (Scheef 2013) performs all aspects of fitting a MAR model to time series data and includes routines to do the critical model search steps. The Supplement shows complete worked examples, including an example of fitting a model with observation error. For fitting MAR state-space models, the MARSS R package is available and the user guide (Holmes et al. 2012a) includes many ecological case studies involving multivariate process models, including lag-p models, with various observation models and a worked example of estimating interactions strengths with error-ridden species and covariate data.

MAR modeling provides a framework for inferring community structure, dynamics, stability and identifying environmental drivers using long-term time series data. With the maturation of many long-term ecological data sets across ecosystems worldwide, MAR modeling is a

Reports

method that should be broadly useful, particularly as new tools make it easier for ecologists to fit these models and modifications accommodate a greater diversity of ecological data sets and questions.

Literature Cited

Beisner, B. E., A. R. Ives, and S. R. Carpenter. 2003. The effects of an exotic fish invasion on the prey communities of two lakes. Journal of Animal Ecology 72:331–342.

Burnham, K. P., and D. R. Anderson. 2010. Model selection and multi-model inference: a practical information-theoretic approach. Springer, New York, New York, USA.

Carpenter, S. R., J. J. Cole, M. L. Pace, M. Van de Bogert, D. L. Bade, D. Bastviken, C. M. Gille, J. R. Hodgson, J. F. Kitchell, and E. S. Kritzberg. 2005. Ecosystem subsidies: terrestrial support of aquatic food webs from ^{13}C addition to contrasting lakes. Ecology 86:2737–2750.

Christensen, V., and C. J. Walters. 2004. Ecopath with Ecosim: methods, capabilities and limitations. Ecological Modelling 172:109–139.

Dalton, M. G. 2001. El Niño, expectations, and fishing effort in Monterey Bay, California. Journal of Environmental Economics and Management 42:336–359.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. Estimating density dependence, process noise, and observation error. Ecological Monographs 76:323–341.

Dhoray, S., and S. S. Teelucksingh. 2007. The implications of ecosystem dynamics for fisheries management: a case study of selected fisheries in the Gulf of Paria, Trinidad. Journal of Environmental Management 85:415–428.

Duffy, M. A. 2007. Selective predation, parasitism, and trophic cascades in a bluegill-*Daphnia*-parasite system. Oecologia 153:453–460.

Durbin, J., and S. J. Koopman. 2012. Time series analysis by state space methods (Oxford Statistical Science Series). Oxford University Press, Oxford, UK.

Fischer, J. M., T. M. Frost, and A. R. Ives. 2001. Compensatory dynamics in zooplankton community responses to acidification: measurement and mechanisms. Ecological Applications 11:1060–1072.

Francis, T. B., M. D. Scheuerell, R. D. Brodeur, P. S. Levin, J. J. Ruzicka, N. Tolimieri, and W. T. Peterson. 2012. Climate shifts the interaction web of a marine plankton community. Global Change Biology 18:2498–2508.

Fulton, E. A., A. D. M. Smith, and C. R. Johnson. 2004. Biogeochemical marine ecosystem models I: IGBEM—a model of marine bay ecosystems. Ecological Modelling 174:267–307.

Grant, R. F. 2001. A review of the Canadian ecosystem model ecosys. Pages 173–264 in M. J. Shaffer, L. Ma, and S. Hansen, editors. Modeling carbon and nitrogen dynamics for soil management. CRC Press, Boca Raton, Florida, USA.

Hairston, N. G., S. P. Ellner, M. A. Geber, T. Yoshida, and J. A. Fox. 2005. Rapid evolution and the convergence of ecological and evolutionary time. Ecology Letters 8:1114–1127.

Hall, S. R., C. R. Becker, J. L. Simonis, M. A. Duffy, A. J. Tessier, and C. E. Caaceres. 2009. Friendly competition: evidence for a dilution effect among competitors in a planktonic host–parasite system. Ecology 90:791–801.

Hampton, S. E., L. R. Izmest'eva, M. V. Moore, S. L. Katz, B. Dennis, and E. A. Silow. 2008. Sixty years of environmental change in the world's largest freshwater lake—Lake Baikal, Siberia. Global Change Biology 14:1947–1958.

Hampton, S. E., M. D. Scheuerell, and D. E. Schindler. 2006. Coalescence in the Lake Washington story: interaction strengths in a planktonic food web. Limnology and Oceanography 51:2042–2051.

Hampton, S. E., and D. E. Schindler. 2006. Empirical evaluation of observation scale effects in community time series. Oikos 113:424–439.

Hinrichsen, R. A., and E. E. Holmes. 2009. Using multivariate state-space models to study spatial structure and dynamics. Pages 1–20 in R. S. Cantrell, C. Cosner, and S. Ruan, editors. Spatial ecology. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Holmes, E. E., J. L. Sabo, S. V. Viscido, and W. F. Fagan. 2007. A statistical approach to quasi-extinction forecasting. Ecology Letters 10:1182–1198.

Holmes, E. E., E. J. Ward, and M. D. Scheuerell. 2012a. Analysis of multivariate time-series using the MARSS package. Version 3.4. http://cran.r-project.org/web/packages/MARSS

Holmes, E. E., E. J. Ward, and K. Wills. 2012b. MARSS: multivariate autoregressive state-space models for analyzing time series data. R Journal 4:11–19.

Huber, V., and U. Gaedke. 2006. The role of predation for seasonal variability patterns among phytoplankton and ciliates. Oikos 114:265–276.

Humbert, J.-Y., L. S. Mills, J. S. Horne, and B. Dennis. 2009. A better way to estimate population trends. Oikos 118:1940–1946.

Ives, A. R. 1995. Predicting the response of populations to environmental change. Ecology 76:926–941.

Ives, A. R., S. R. Carpenter, and B. Dennis. 1999. Community interaction webs and zooplankton responses to planktivory manipulations. Ecology 80:1405–1421.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs 73:301–330.

Klug, J. L., and K. L. Cottingham. 2001. Interactions among environmental drivers: community responses to changing nutrients and dissolved organic carbon. Ecology 82:3390–3403.

Klug, J. L., J. M. Fischer, A. R. Ives, and B. Dennis. 2000. Compensatory dynamics in planktonic community responses to pH perturbations. Ecology 81:387–398.

Knape, J., and P. de Valpine. 2012. Are patterns of density dependence in the global population dynamics database driven by uncertainty about population abundance? Ecology Letters 15:17–23.

Kordas, R. L., and S. Dudgeon. 2011. Dynamics of species interaction strength in space, time and with developmental stage. Proceedings of the Royal Society B 278:1804–1813.

Lindegren, M., C. Mollmann, A. Nielsen, and N. C. Stenseth. 2009. Preventing the collapse of the Baltic cod stock through an ecosystem-based management approach. Proceedings of the National Academy of Sciences USA 106:14722–14727.

Mac Nally, R., et al. 2010. Analysis of pelagic species decline in the upper San Francisco Estuary using multivariate autoregressive modeling (MAR). Ecological Applications 20:1417–1430.

Mutshinda, C. M., R. B. O'Hara, and I. P. Woiwod. 2009. What drives community dynamics? Proceedings of the Royal Society of London B 276:2923–2929.

Neubert, M. G., H. Caswell, and A. R. Solow. 2009. Detecting reactivity. Ecology 90:2683–2688.

Plagányi, E. E. 2007. Models for an ecosystem approach to fisheries. FAO Fisheries Technical Paper 477. FAO, Rome, Italy.

Scheef, L. 2013. MAR1: multivariate autoregressive modeling for analysis of community time-series data. R package

Reports

version 1.0. http://cran.r-project.org/web/packages/MAR1/index.html

Scheef, L. P., S. E. Hampton, and L. R. Izmest'eva. *In press.* Inferring plankton community structure from marine and freshwater long-term data using multivariate autoregressive models. Limnology and Oceanography: Methods.

Scheef, L. P., D. E. Pendleton, S. E. Hampton, S. L. Katz, E. E. Holmes, M. D. Scheuerell, and D. G. Johns. 2012. Assessing marine plankton community structure from long-term monitoring data with multivariate autoregressive (MAR) models: a comparison of fixed station versus spatially distributed sampling data. Limnology and Oceanography: Methods 10:54–64.

Shenk, T. M., G. C. White, and K. P. Burnham. 1998. Sampling-variance effects on detecting density dependence from temporal trends in natural populations. Ecological Monographs 68:445–463.

Shin, Y. J., and P. Cury. 2004. Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. Canadian Journal of Fisheries and Aquatic Sciences 61:414–431.

Shumway, R. H., and D. S. Stoffer. 2006. Time series analysis and its applications with R examples. Springer, New York, New York, USA.

Stelzer, C.-P. 1998. Feeding behaviour of the rotifer *Ascomorpha ovalis:* functional response, handling time and exploita-

tion of individual *Ceratium* cells. Journal of Plankton Research 20:1131–1144.

Stergiou, K. I. 1991. Describing and forecasting the sardine–anchovy complex in the eastern Mediterranean using vector autoregressions. Fisheries Research 11:127–141.

Stergiou, K., and E. Christou. 1996. Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods. Fisheries Research 25:105–138.

Thébault, E., and C. Fontaine. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853–856.

Thibaut, L. M., S. R. Connolly, and H. P. A. Sweatman. 2012. Diversity and stability of herbivorous fishes on coral reefs. Ecology 93:891–901.

Tsay, R. S. 2010. Analysis of financial time series. Wiley, Cambridge, Massachusetts, USA.

Vik, J. O., C. N. Brinch, S. Boutin, and N. C. Stenseth. 2008. Interlinking hare and lynx dynamics using a century's worth of annual data. Population Ecology 50:267–274.

Yamamura, K., M. Yokozawa, M. Nishimori, Y. Ueda, and T. Yokosuka. 2006. How to analyze long-term insect population dynamics under climate change: 50-year data of three insect pests in paddy fields. Population Ecology 48:31–48.

SUPPLEMENTAL MATERIAL

Supplement

Applying MAR(1) and MARSS to an ecological data set (*Ecological Archives* E094-247-S1).

Reports

Applying MAR(1) and MARSS to an ecological dataset
Supplement to "Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models" (S.E. Hampton, E.E. Holmes, L.P. Scheef, M.D. Scheuerell, S.L. Katz, D.E. Pendleton, E.J Ward, submitted to Ecology18 July 2013)

Here we step through an example of applying both MAR(1) and MARSS to an ecological dataset, using R, a free open-source software environment for statistical computing and graphics (http://www.r-project.org/). The example R code is inset in the text, but also available as a block at the end of this supplement for those who would prefer to cut and paste the whole block into an R text editor before beginning the exercise. At the time of publication, this exercise had been tested in R versions 2.15.1 and 3.0.1, primarily in 64-bit installations.

MAR1 R package

The tools necessary to apply a basic MAR(1) model to time-series data have been built into the MAR1 package for R (Scheef 2013, http://cran.r-project.org/web/packages/MAR1/). This package uses the tcltk package to create pop-up windows. On Mac OS X systems, installation of the Tcl/Tk 8.5.5 library for X11 may be necessary before the Tcl/Tk interface can successfully load. See http://cran.r-project.org/bin/macosx/tools/ for details. It is possible to run the package without the pop-up windows.

From within the R console, the MAR1 package can be installed by typing or pasting in:

```
install.packages("MAR1", repos="http://cran.rstudio.com",
    dependencies=TRUE)
```

To load the package in the current R session:

```
library(MAR1)
```

Prepare the dataset

The MAR1 package includes weekly marine plankton abundance data from the Western Channel Observatory (WCO) L4 station (www.pml.ac.uk/L4) as an example of a plankton time-series that is of appropriate temporal and taxonomic resolution for MAR application. See (2010) and (Southward et al. 2005) for L4 zooplankton and phytoplankton sample collection and processing methods. The L4 data subset included in the MAR1 package is a weekly time-series of sea surface temperature values (°C) and abundances for 15 zooplankton groups (individuals m^{-3}) and 3 phytoplankton groups (cells ml^{-1}). The plankton groups were created by summing the abundance values of taxonomically and ecologically similar taxa included in the raw L4 time-series for each time-step.

To load the L4 time-series, take a look at the column heads and the first few lines of data, see a statistical summary of the dataset, and see additional details about the dataset in its associated help file:

```
data(L4.AllDates)
head(L4.AllDates)
summary(L4.AllDates)

?L4.AllDates
```

Creating a time-series with evenly spaced intervals

We can see that the L4 dataset has a format common to time-series data: sampling dates are recorded in the first column and the measured values of each parameter are recorded in the following columns. Columns containing information other than dates and the values of potential variables have been removed to simplify the dataset for MAR application. Because zooplankton and phytoplankton were not always sampled on the same date, we can see in the summary that the data include many missing (NA) values.

For MAR, the data need to be averaged into evenly-spaced time-steps. The MAR1 package includes a function for transforming datasets that can average data into monthly, yearly, weekly, or daily time-steps. To apply this function, the dataset must be arranged so that dates are in the first column and the numeric values of variables are in the following columns. Dates must include four-digit year values (yyyy) and some character separating year, month, and day (e.g., 1982-07-26 or 2003/11/06). The order that year, month, and day are represented in the date values does not matter. The L4 dataset we are using conforms to these requirements, so the function can be directly applied. Monthly increments have been reported as an appropriate time-scale for the detection of plankton community interactions in other MAR studies (e.g., Hampton and Schindler 2006), so we will average the L4 data into monthly time-steps.

To average the data into monthly increments and view the summary and dimensions (number of rows and columns) of the new time-series object:

```
L4.byMonth <- prepare.data(data=L4.AllDates,
     increment="month")

summary(L4.byMonth)
dim(L4.byMonth)
```

We see that the data have been averaged into 175 continuous time-steps, and all NA values have been eliminated in the process. Note that 175 is not the total number of time-steps in the dataset (which is 179 according to the dataset dimensions), but the number of time-steps that occur without an adjacent missing value in the series. This is the number of data points that will be used to build the model, as non-consecutive time-points will be thrown out during the analysis. A new column 'contin' has also been created, in which each run of equal values designates a continuous block of time-steps.

Data transformations

In plankton MAR studies, linear interpolation has sometimes been used to fill in small gaps in the time-series. However, when data consistently miss a certain time period (e.g., no winter sampling) investigators frequently just exclude those time periods from the time-series (e.g., Ives et al. 2003, Hampton et al. 2006) and allow MAR to skip estimations between non-consecutive data points. It is also common practice to log-transform data in order to better approximate the non-linear relationships frequently present in ecological data (Ives et al. 2003). If the data are log-transformed, zeros need to be replaced with non-zero values. For some questions investigators standardize the data to dimensionless units (Z-scores) so that model results can be directly compared among plankton groups (e.g., Hampton et al. 2006). The same function used to average the data into evenly spaced time-steps can perform these transformations as well.

To see all arguments for the `prepare.data` function and possible values for each argument, and to see the help file for the function:

```
formals(prepare.data)

?prepare.data
```

If we decide to fill gaps in our time-series via linear interpolation, we can choose how large a gap we will allow to be filled with `fill.gap`. For example, if we set `fill.gap=2`, gaps two months long or less will be interpolated, and larger gaps will be left in the time-series.

If we would like to log10-transform the data (by setting `log=TRUE`), zeros in the time-series must first be replaced. Zero values occurring in the time-series for a particular plankton group can be replaced with a random number between 0 and half the lowest non-zero value for that group (`replace.0s="rand.half"`), or zeros can be eliminated by adding 1 to all values (`replace.0s="add.ones"`).

There are two methods that can be used to Z-score the data as well. The first method (`z.method="standard"`) standardizes each value by subtracting the overall mean abundance value for the group and dividing by the overall standard deviation for the group. The second method (`z.method="deseason"`) is similar, except the means and standard deviations are month-specific so that mean seasonal trends are removed from the data-series. De-seasoning the abundance data could aid in the detection of interactions between plankton groups by dampening seasonal successions that relate to seasonally varying abiotic drivers.

We will apply the `prepare.data` function to the L4 data again, but in addition to averaging the data into monthly increments, we will transform the data so that gaps of one month are filled via linear interpolation, zeros are replaced with a random number between 0 and half the lowest value for the respective group, and the values are log10-transformed and then Z-scored using the "standard" method:

```
L4.mar1 <- prepare.data(data=L4.AllDates,
    increment="month", fill.gap=1, replace.0s="rand.half",
    log=T, z.method="standard")
```

R reports that two time gaps were filled, that the time-series now contains 179 continuous time-steps, and that the log and Z-score transformations were successful. The function that runs the MAR model requires that the data be arranged with the continuous time-block variable in the first column, dates in the second column, and the time-series for the variables in the following columns. Looking at the summary of the transformed dataset, we see that it is appropriately formatted for MAR application.

```
summary(L4.mar1)
```

To see how the Z-score method used affects the MAR results, we can create a second transformed dataset with the "deseason" Z-score method for comparison:

```
L4.mar2 <- prepare.data(data=L4.AllDates,
    increment="month", fill.gap=1, replace.0s="rand.half",
    log=T, z.method="deseason")

summary(L4.mar2)
```

Build the MAR(1) model

To see the arguments for the function that builds the MAR model and to see the help file for the function:

```
formals(run.mar)

?run.mar
```

The function `run.mar` allows the user to select which interactions will be included in the model. The variables argument designates whether variables will be included as variates (X in Equation 3) or covariates (U in Equation 3). The user can also set restrictions on interactions between the variables: interactions can be forced to be included in or excluded from the final model (the `restrictions` argument). By default, a random search for the best-fit model is performed (`search="random"`) following Ives et al. (2003), but options for forward step and exhaustive searches are also available (see `?run.mar`). The function does a 500 iteration bootstrap of the best-fit model coefficients by default, and the number of iterations can be changed (`boot=`*n*), or bootstrapping can be skipped altogether (`boot=FALSE`). The top 10 models (i.e., lowest AIC models, including the best-fit model) are retained in the output by default, and this number also can be changed (`ntop=`*n*) or excluded (`ntop=FALSE`). All results of the MAR analysis are stored as a list object within R (see `?run.mar` for details on the structure and components of the list object). If the `export` argument is set to `TRUE` or to a quoted name (e.g., `export="FolderName"`), a folder will be created in the current working directory (which can be viewed with `getwd()`), and all components of the result list will be saved as comma-delimited (.csv) files in that folder. A MAR result list object can also be exported later using the function `export.MAR()` (see `?export.MAR`).

For the L4 data, one might be interested in using MAR to assess whether there are detectible interactions between invertebrate predators and copepods. It is thought that cnidarians and chaetognaths may have important predatory effects on calanoid copepods in the English Channel (Irigoien and Harris 2003, Eloire et al. 2010, Bonnet et al. 2010). To investigate this hypothesis with MAR, we will build a model that includes the cnidarian, chaetognath, large calanoid, and small calanoid zooplankton groups and the diatom and dinoflagellate phytoplankton groups as variates, and sea surface temperature as a covariate.

To initiate the analysis:

```
run1 <- run.mar(data=L4.mar1)
```

Because no values were provided for the `variables` argument, a window that allows the user to select variates and covariates with toggle buttons appears, along with instructions in the R Console. Click the buttons next to <u>cnidarian</u>, <u>chaetognath</u>, <u>calanoid.lg</u>, <u>calanoid.sm</u>, <u>diatom</u>, and <u>dino</u> once to select them as variates, and the button next to <u>surface.temp</u> twice to select it as a covariate. Clicking a button three times resets it to 'not included.' After clicking Done, another window appears to allow restrictions to be set on interactions with toggle buttons. To see what the MAR model will look like when we don't set any restrictions, click Done. A progress report shows in the console while the function is working, and all results are stored in the 'run1' object.

The analysis can also be run without using the windows to select the variables and set restrictions. The variables argument can be set equal to a vector in which each column in the dataset is represented by a 0 (not included), 1 (variate), or 2 (covariate). The first two numbers of the vector represent the 'contin' and time-step columns and should always be 0. The restrictions argument can be set equal to a matrix containing a row for every variate and a column for every variate and covariate, where the interaction of each column term on each row variate is represented by a 0.5 (may be included in the model), 0 (forced exclusion from the model), or 1 (forced inclusion in the model). To build the 'run1' model above without using the pop-up windows:

```
myvar <- c(0,0,1,0,1,0,0,0,0,1,1,0,0,0,1,1,0,0,0,0,2)
myres <- matrix(0.5, nrow=length(which(myvar==1)),
     ncol=length(which(myvar!=0)))

run1 <- run.mar(data=L4.mar1, variables=myvar,
     restrictions=myres)
```

To print the main results, see a summary of the results, and plot the model:

```
run1
summary(run1)
plot(run1)
```

In the plot of interaction strengths, bars extending to the right and left of the dotted lines represent positive and negative interactions, respectively. Interactions in the best-fit model that were excluded by bootstrapping are plotted as lighter, hatched bars. It is apparent that the analysis detected mostly positive interactions. All variates exhibit positive density dependence along the diagonal of the B-matrix, and large calanoid copepods and diatoms appear to have positive effects on the abundances of nearly all the other variates. Surface temperature, the only covariate included in the model, is shown to have negative effects on diatoms and large and small calanoids. Several of the interactions, including the positive effects of large calanoids on both phytoplankton groups and the negative effects of temperature on the variates, seem ecologically improbable. To assess whether the interaction patterns in this model might have been influenced by the detection of seasonal successions in the time-series, we can build another model with our "deseason" Z-scored dataset:

```
run2 <- run.mar(data=L4.mar2, variables=run1,
     restrictions=run1)
```

Since we are using a dataset with identical formatting to the one used in building the first model (i.e., same column order and column headers) and we would like to use the same variables and restrictions as those set in the first model, we can simply set the variables and restrictions arguments of the function equal to the 'run1' object.

To compare both models in one plot and create a legend for the plot:

```
plot(run1, run2, legend=T)
```

We can see that the second model for the de-seasoned data includes notably fewer interactions than the first model. None of the negative temperature effects were detected in the second model, and the only positive effect of large calanoids on any variate is their own density dependence. The second model also includes a plausible negative effect of large calanoids on diatoms that was not detected in the first model. An unlikely positive effect of cnidarians on chaetognaths was detected in both models. If we wanted to recreate the second, de-seasoned data model, but with the interaction of cnidarians on chaetognaths excluded, we can set that restriction by toggling the corresponding button (first column, second row) to 0 in the Specify Restrictions window before clicking Done:

```
run3 <- run.mar(data=L4.mar2, variables=run2)
```

Or, to do the same without use of the pop-up window:

```
myres[2,1] <- 0

run3 <- run.mar(data=L4.mar2, variables=run2,
     restrictions=myres)
```

To plot all three models:

```
plot(run1, run2, run3, legend=T)
```

In the plot, it is apparent that setting this cnidarian-chaetognath restriction (indicated by a red dot) had little effect on the results.

To compare the best-fit model to the other top models retained in the analysis, we can plot the '$top.bestfit' component of the result list object:

```
dev.new()
plot(run3$top.bestfit)
```

Here, the best-fit model is represented by the top-most set of bars, and nine other low-AIC models are plotted below it in order of increasing AIC.

To view a histogram of the AIC values for the best models:

```
hist(run3$top.bestfit)
```

The AIC value for the selected best-fit model is denoted by the blue star along the bottom of the plot.

Building a state-space MAR model (MARSS)

A state-space MAR model is a MAR(1) model with an observation process:

$$\mathbf{X}_t = \mathbf{A} + \mathbf{B}\mathbf{X}_{t-1} + \mathbf{C}\mathbf{U}_t + \mathbf{E}_t$$
$$\mathbf{Y}_t = \mathbf{X}_t + \mathbf{V}_t$$

S1

The underlying community dynamics are described by the MAR(1) model ($\mathbf{X}$) but the true values of $\mathbf{X}$ are "hidden", and we only observe $\mathbf{Y}_t$ which is $\mathbf{X}_t$ plus some error. State-space MAR (MARSS) might be particularly helpful with this dataset, since we anticipate that the observation error is a real issue in this relatively open marine system.

The MARSS package (Holmes et al. 2013, http://cran.r-project.org/web/packages/MARSS/) is available in R for fitting state-space multivariate autoregressive models, and we can use it to fit a state-space version of the MAR(1) models in the examples above.

To install and load the MARSS package:

```
install.packages("MARSS", repos="http://cran.rstudio.com",
    dependencies=TRUE)

library(MARSS)
```

The MAR1 package has a function `ss.mar1()` which will allow you to easily fit the model described in eqn. S1: our observations are $\mathbf{X}_t$ plus error and we will treat our covariates $\mathbf{U}_t$, in this case surface temperature, as known without observation error. Although in this example we use the simple observation process described by eqn. S2, the MARSS package will allow you to model much more general observation processes and can allow one to model, for example, situations with multiple measurement sites, measurement equipment or sampling designs that change over time, sampling sites with different levels of missing values, or covariates that are measured with error and in multiple locations. See the MARSS User Guide on the MARSS CRAN webpage (http://cran.r-project.org/web/packages/MARSS) for a discussion of analysis of species interactions using more complicated observation models.

The `ss.mar1()` function uses an aggregated dataset as input. This is a dataset already transformed by `prepare.data()`. It also needs to have the MAR model $\mathbf{B}$ and $\mathbf{C}$ matrices specified. One way to do this is to pass in a MAR fitted model object generated by `run.mar()`, and `ss.mar1()` will use the bestfit B and C.

Here we pass the de-seasoned L4 dataset and the third MAR model we generated above to the state-space function. (Note: This may take a few moments to run. On 32-bit R installations, error messages sometimes occur in this step though results are comparable; if you receive errors, to generate your final results you may want to experiment with a different version or installation.)

```
ss.run3 <- ss.mar1(L4.mar2, MAR.obj=run3)
```

We can compare the **B** matrices from the non-state-space MAR and the state-space MAR by printing them and/or plotting them:

```
run3$bestfit$B
ss.run3$B

ss.run3.plot <-
      list(restrictions.set=run1$restrictions.set,
      bestfit=list(B=ss.run3$B,C=ss.run3$C), bootstrap=NULL)
class(ss.run3.plot) <- "MAR"

dev.new()
plot(run3, ss.run3.plot, legend=T)
```

The first thing that is apparent is that the interaction strengths along the diagonal are larger, potentially indicating that the observation error is now being more appropriately accounted for. Observation error tends to make time-series look like white-noise and thus moves the diagonal terms towards 0. We also see that the off-diagonal terms are farther from 0 with the state-space MAR model. This is also suggestive of better partitioning of the variance. Observation error tends to make species look less correlated and thus their interactions strengths are closer to 0. Overall however, the direction of the interactions are largely concordant with the results from the MAR(1) model.

The L4.mar2 dataset had gaps of 1 filled by linear interpolation. A state-space MAR model can handle missing values fine because it incorporates an observation model.

We can compare the **B** estimates if we do not do linear interpolation.

```
# L4 data aggregated by month with gaps not filled

L4.mar3 <- prepare.data(data=L4.AllDates,
      increment="month", fill.gap=0, replace.0s="rand.half",
      log=TRUE, z.method="deseason")

ss.run3.no.interp <- ss.mar1(L4.mar3, MAR.obj=run3)

ss.run3.no.interp$B
```

These models take a long time to fit because estimating both the observation and process variance matrices is difficult. We can also fix the observation variance by passing an explicit V_t matrix into `ss.mar1()`. The observation variance matrix is termed R, and we will set the observation variance of the zooplankton to be 0.1 and the dinoflagellates and diatoms to be 0.05. The model now fits much faster.

```
# Set observation variance matrix instead of estimating it

R <- diag(c(0.1,0.1,0.1,0.1,0.05,0.05))

ss.run3.fixed.R <- ss.mar1(L4.mar3, MAR.obj=run3,
```

```
        model=list(R=R))

ss.run3.fixed.R$B
```

To compare the non-state-space and all three state-space MAR models in a plot:

```
ss.run3.no.interp.plot <-
     list(restrictions.set=run1$restrictions.set,
     bestfit=list(B=ss.run3.no.interp$B,C=ss.run3.no.interp
     $C), bootstrap=NULL)
class(ss.run3.no.interp.plot) <- "MAR"

ss.run3.fixed.R.plot <-
     list(restrictions.set=run1$restrictions.set,
     bestfit=list(B=ss.run3.fixed.R$B,C=ss.run3.fixed.R$C),
     bootstrap=NULL)
class(ss.run3.fixed.R.plot) <- "MAR"

dev.new()
plot(run3, ss.run3.plot, ss.run3.no.interp.plot,
     ss.run3.fixed.R.plot, legend=T)
```

Code from this exercise

Some readers may find it easiest to just cut and paste this whole block into an R text editor before working through the exercise.

```
######## MAR1 R Package ########

# Install MAR1 package

install.packages("MAR1", repos="http://cran.rstudio.com",
     dependencies=TRUE)

# Load package

library(MAR1)

######## Prepare the dataset ########

# Load and view L4 dataset included in MAR1 package

data(L4.AllDates)
head(L4.AllDates)
summary(L4.AllDates)
```

```
?L4.AllDates

# Begin transformations - average into monthly time steps

L4.byMonth <- prepare.data(data=L4.AllDates,
    increment="month")

summary(L4.byMonth)
dim(L4.byMonth)

# See all arguments and help file for prepare.data function

formals(prepare.data)
?prepare.data

# Fill gaps with linear interpolation, perform log and z-
    score transformations

L4.mar1 <- prepare.data(data=L4.AllDates,
    increment="month", fill.gap=1, replace.0s="rand.half",
    log=T, z.method="standard")

# Look at summary of transformed dataset

summary(L4.mar1)

# Create another dataset using the de-seasoning z-score
    method

L4.mar2 <- prepare.data(data=L4.AllDates,
    increment="month", fill.gap=1, replace.0s="rand.half",
    log=T, z.method="deseason")

summary(L4.mar2)

######## Build the MAR(1) model ########

# Look at the arguments available in run.mar and view help
    file

formals(run.mar)
?run.mar

# Initiate MAR(1) analysis using pop-up windows to select
    variables and set interaction restrictions
    # Set cnidarian, chaetognath, calanoid.lg,
    calanoid.sm, diatom, and dino as variates and
    surface.temp as a covariate
    # Do not set any restrictions

run1 <- run.mar(data=L4.mar1)
```

```
# Same model can be built without using the pop-up windows

myvar <- c(0,0,1,0,1,0,0,0,0,1,1,0,0,0,1,1,0,0,0,0,2)
myres <- matrix(0.5, nrow=length(which(myvar==1)),
      ncol=length(which(myvar!=0)))

run1 <- run.mar(data=L4.mar1, variables=myvar,
      restrictions=myres)

# Summarize and plot main results

run1
summary(run1)
plot(run1)

# Build another model using the de-seasoned dataset

run2<-run.mar(data=L4.mar2, variables=run1,
      restrictions=run1)

# Compare results of 1st and 2nd model in a plot

plot(run1, run2, legend=T)

# Run a 3rd model - this time you can toggle off one of the
      biologically implausible interactions (cnidarians on
      chaetognaths)

run3 <- run.mar(data=L4.mar2, variables=run2)

# Same model can be built without using the pop-up window

myres[2,1]<-0

run3 <- run.mar(data=L4.mar2, variables=run2,
      restrictions=myres)

# Compare results of all three models in a plot

plot(run1, run2, run3, legend=T)

# For top model (lowest AIC) from 3rd model, compare to
      other top models in a plot

dev.new()
plot(run3$top.bestfit)

# Compare AIC values of top models using a histogram

hist(run3$top.bestfit)
```

```
######## Building a state-space MAR model (MARSS) ########

# Install and load the MARSS package

install.packages("MARSS", repos="http://cran.rstudio.com",
      dependencies=TRUE)

library(MARSS)

# Compare non-state-space MAR to a state-space MAR

ss.run3 <- ss.mar1(L4.mar2, MAR.obj=run3)

ss.run3$B
run3$bestfit$B

ss.run3.plot <-
      list(restrictions.set=run1$restrictions.set,
      bestfit=list(B=ss.run3$B,C=ss.run3$C), bootstrap=NULL)
class(ss.run3.plot) <- "MAR"

dev.new()
plot(run3, ss.run3.plot, legend=T)

# Use a state-space model with L4 data aggregated by month
      with gaps not filled

L4.mar3 <- prepare.data(data=L4.AllDates,
      increment="month", fill.gap=0, replace.0s="rand.half",
      log=TRUE, z.method="deseason")

ss.run3.no.interp <- ss.mar1(L4.mar3, MAR.obj=run3)

ss.run3.no.interp$B

# Use a state-space model with a fixed observation variance

R <- diag(c(.1,.1,.1,.1,.05,.05))

ss.run3.fixed.R <- ss.mar1(L4.mar3, MAR.obj=run3,
      model=list(R=R))

ss.run3.fixed.R$B

# Compare non-state-space and all three state-space MAR
      models in a plot

ss.run3.no.interp.plot <-
      list(restrictions.set=run1$restrictions.set,
```

```
        bestfit=list(B=ss.run3.no.interp$B,C=ss.run3.no.interp
        $C), bootstrap=NULL)
class(ss.run3.no.interp.plot) <- "MAR"

ss.run3.fixed.R.plot <-
        list(restrictions.set=run1$restrictions.set,
        bestfit=list(B=ss.run3.fixed.R$B,C=ss.run3.fixed.R$C),
        bootstrap=NULL)
class(ss.run3.fixed.R.plot) <- "MAR"

dev.new()
plot(run3, ss.run3.plot, ss.run3.no.interp.plot,
        ss.run3.fixed.R.plot, legend=T)
```

References

Bonnet, D., P. K. Lindeque, and R. P. Harris. 2010. Sagitta setosa predation on Calanus helgolandicus in the English Channel. Journal of Plankton Research 32:725–737.

Eloire, D., P. J. Somerfield, D. V. P. Conway, C. Halsband-Lenk, R. Harris, and D. Bonnet. 2010. Temporal variability and community composition of zooplankton at station L4 in the Western Channel: 20 years of sampling. Journal of Plankton Research 32:657–679.

Hampton, S. E., M. D. Scheuerell, and D. E. Schindler. 2006. Coalescence in the Lake Washington story: interaction strengths in a planktonic food web. Limnology and Oceanography 51:2042–2051.

Hampton, S. E., and D. E. Schindler. 2006. Empirical evaluation of observation scale effects in community time series. Oikos 113:424–439.

Holmes, E., E. Ward, and K. Wills. 2013. MARSS: Multivariate autoregressive state-space modeling. R package version 3.4.

Irigoien, X., and R. P. Harris. 2003. Interannual variability of Calanus helgolandicus in the English Channel. Fisheries Oceanography 12:317–326.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs 73:301–330.

Scheef, L. 2013. MAR1: Multivariate Autoregressive Modeling for Analysis of Community Time-Series Data. R package version 1.0.

Southward, A. J., O. Langmead, N. J. Hardman-Mountford, J. Aiken, G. T. Boalch, P. R. Dando, M. J. Genner, I. Joint, M. A. Kendall, N. C. Halliday, R. P. Harris, R. Leaper, N. Mieszkowska, R. D. Pingree, A. J. Richardson, D. W. Sims, T. Smith, A. W. Walne, and S. J. Hawkins. 2005. Long-term oceanographic and ecological research in the western English Channel. Advances in Marine Biology 47:1–105.