Inferring plankton community structure from marine and freshwater long-term data using multivariate autoregressive models

Lindsay P. Scheef[1]*, Stephanie E. Hampton[1], and Lyubov R. Izmest'eva[2]
[1]National Center for Ecological Analysis & Synthesis, University of California, Santa Barbara, CA 93101, USA
[2]Scientific Research Institute of Biology, Irkutsk State University, Irkutsk, 664003, Russia

Abstract

Multivariate autoregressive (MAR) models have been useful in elucidating food web dynamics and stability from freshwater plankton monitoring data, but their applicability to marine datasets has not been as well explored. Characteristics of marine systems, such as the movement of water masses by tides and currents, may present unique challenges to MAR modeling of data gathered in marine environments. To explore the behavior of MAR models with marine plankton data, in the context of what we know about applying MAR to freshwater data, we applied MARs to each of three freshwater and four marine long-term datasets and compared results among them. We generated sets of replicate MAR models for each dataset and used the consistency of models within each set of replicates as a measure of MAR performance. Overall, replicate MAR models generated from the marine datasets were less consistent than those generated from the freshwater datasets, suggesting that MAR methods need fundamental reconfigurations to be applied to standard marine plankton data. Higher variability observed within the marine MAR results may be attributable to weaker biotic interactions as represented by the data, and to overparameterization when the criteria for lumping freshwater plankton taxa into model variables are directly applied to marine plankton taxa. Adjustments to dataset preparation for MAR application and to the modeling framework itself may address these issues associated with analyzing data from highly dynamic systems.

A central challenge in managing natural resources lies in anticipating the ecosystem consequences of human actions that resonate through food webs via direct and indirect interactions (Pace et al. 1999; Soulé et al. 2005). Even the most general predictions for trophic repercussions of human perturbations require some basic understanding of the food web at hand. Fundamental descriptions of food webs have historically relied on a mix of direct observation and targeted experimentation with communities of interest (Paine 1980), and numerous recent advances in stable isotope analysis have allowed relatively coarse characterizations of food webs (Cabana and Rasmussen 1994; Layman et al. 2007). However, in addition to identifying the realized links among taxa in a food web, it is critically important to understand the strength of these interactions to anticipate ecosystem responses that emerge when particular taxa are perturbed (Soulé et al. 2005; Bascompte et al. 2005; Ings et al. 2009). Unfortunately, the detailed knowledge of species interactions that would allow for the construction of a food web, much less the quantification of interaction strengths, is lacking for many ecosystems. In some of these systems, however, long-term monitoring data on population abundances are available, and multivariate analysis of these time-series may give a unique perspective on how trophic relationships and interaction strengths, as well as the abiotic drivers, affect the communities.

Multivariate autoregressive (MAR) modeling estimates the strengths of interactions among species and environmental drivers from time-series data (Ives 1995; Ives et al. 1999, 2003). Unlike most statistical approaches familiar to ecologists, MARs use the temporal autocorrelation in the data to detect time-

lagged population responses, rather than treating autocorrelation as a nuisance. For example, in the MAR approach, an increase in a predator group may be correlated with a decrease in prey populations in the following time step. Variance in population abundance is partitioned among potentially interacting biotic and abiotic variables, such that direct and indirect pathways within food webs can be discerned. The MAR approach can helpfully identify the strongest interactions in a food web, eliminating the species interactions that are too weak to be useful predictors of population dynamics. As in all statistical analyses of ecological data, the resulting food web characterization is only as robust as the investigators' knowledge of the ecosystem's natural history.

The MAR approach may not be appropriate for every dataset. Suitability of MAR for application to any given dataset depends to a large degree on the sampling regime used and the length of the time-series, as well as the characteristics of the environment and biotic community sampled. Sampling methods and frequency affect the magnitude and form of error included in the data, and some environments are more difficult than others to fully characterize in a monitoring program, leaving a relatively large number of environmental drivers unmeasured or important spatial processes poorly accounted for in population dynamics (e.g., upwelling in coastal oceans). Also, some communities may be comprised of organisms with a broad diversity of life histories that greatly complicate the application and interpretation of MAR analysis. For example, MAR becomes increasingly difficult for communities in which key players have distinctive ontogenetic niche shifts, exhibit strong migration patterns, or respond to each other on radically different temporal scales (e.g., whales and plankton). Even if these challenges are minimal, the effectiveness of a MAR model in reflecting the true mean state of a system is dependent on the number of time-steps included in the analysis, so MAR is expected to perform best on relatively long, multi-year time-series.

Several studies have demonstrated that plankton time-series data collected from lakes can be analyzed with MAR to gain insights on freshwater plankton community dynamics (e.g., Ives et al. 1999; Hampton and Schindler 2006; Huber and Gaedke 2006; Hampton et al. 2008). The MAR models generated from these datasets provide critical clues to how various drivers can influence plankton communities, and in turn, impact higher trophic levels and overall ecosystem health. The results of these studies indicate MAR analyses work well with long plankton time-series (e.g., > 5 y) of fine temporal resolution (e.g., weekly sampling) and small spatial scale (single fixed point) that have been averaged into biweekly to monthly time-steps.

Plankton time-series with these characteristics are also available for marine systems. MAR analysis of these datasets could yield valuable information on how perturbations, such as nutrient loading and climate change, are affecting these ecosystems and the services they provide. However, marine systems present several potential challenges to MAR analysis relative to freshwater lakes. Marine systems are more open, can be strongly influenced by unmeasured environmental drivers (e.g., circulation, wave energy, upwelling, and benthic-pelagic processes), and tend to be sampled less frequently and more irregularly at multiple sampling points. Because some of these issues are likely to be common among a variety of other ecosystem types and methods for successful MAR use with plankton data have already been established, an investigation of how MAR analysis behaves when applied to marine plankton time-series is a logical first step toward adapting the model framework for broader use in ecology.

To investigate the behavior of MAR analysis with marine datasets, we can assess how well the results match our natural history knowledge and evaluate statistical measures of model fit, such as R^2. Because we know MAR is "well-behaved" with certain freshwater plankton datasets (i.e., it has performed consistently in multiple independent studies), we can also compare "marine MARs" to "freshwater MARs" to get a sense for how the analysis might perform differently between the two different types of data.

One interesting characteristic of MAR analysis that may be a useful indicator of its performance is that some of the routines within the analysis, such as the application of a bootstrap to coefficient estimates, include randomness that can allow slightly different results to be generated when the analysis is repeated on the same dataset. Ideally, the models generated by multiple runs of the analysis on a given dataset will be qualitatively identical—the same interactions are definitively detected each time. If the results are ecologically sensible and if other metrics of model performance, such as R^2 values, suggest a good model fit (see Ives et al. 2003), one can feel confident that the analysis is performing reliably and is appropriate for the dataset. On the other hand, if different models are generated each time the analysis is run, one might get an impression that inferences based on any one particular model will be unreliable. If highly inconsistent MAR results are generated despite careful data preparation, it is likely the model framework is not appropriate for the dataset.

In this study, we use the consistency of models generated from multiple model runs as a metric to compare the performance of MAR between freshwater and marine plankton datasets. We expected more consistent MAR results from freshwater time-series, due to the likelihood that the more open and dynamic environmental conditions in marine systems would lead to higher error and lower temporal autocorrelation in marine data.

Materials and procedures

Dataset selection

Data from two freshwater lakes (Lake Baikal, Siberia and Lake Washington, USA) and two marine systems (the English Channel, UK and the Chesapeake Bay, USA) were included in this study. Each of these sampling locations is described below,

Table 1. Characteristics of each freshwater and marine dataset analyzed with MAR.

System type	Dataset	Abbreviation	Description	Temporal range
Freshwater	Lake Baikal	LB	Subset of longer time-series	1974–1989
Freshwater	Lake Washington, part 1	LW$_1$	Prior to *Daphnia* establishment	1962–1977
Freshwater	Lake Washington, part 2	LW$_2$	After *Daphnia* establishment	1976–1991
Marine	English Channel	EC	Western Channel Observatory station L4	1992–2007
Marine	Chesapeake Bay, lower	CB$_L$	Station CB6.4; mean salinity ~ 21	1985–1999
Marine	Chesapeake Bay, middle	CB$_M$	Station CB4.3C; mean salinity ~ 16	1984–2000
Marine	Chesapeake Bay, upper	CB$_U$	Station CB2.2; mean salinity ~ 3	1984–2002

and characteristics of each dataset selected for analysis are given in Table 1.

Lake Baikal

Lake Baikal (abbreviated LB), the world's deepest, highest volume freshwater lake, has been monitored by three generations of biologists within a single Siberian family since 1945 (Kozhova and Izmest'eva 1998; Izmest'eva 2006). Zooplankton and phytoplankton samples are taken at 7-10 days intervals throughout the year and enumerated at the species level. A detailed description of the Lake Baikal sample collection and processing methods is given by Hampton et al. (2008). An available subset of the entire time-series (1974-1989) was selected for inclusion in this study.

Lake Washington

Lake Washington (LW) has been monitored on a 1-4 week basis since 1962. Data from surface collections at a mid-lake station (Madison Park) represent the most complete time-series for the lake and were selected for analysis (Hampton et al. 2006). Edmondson and Lehman (1981), Edmondson and Litt (1982), and Edmondson et al. (2003) provide details for Lake Washington sample collection and processing methods. The time-series was divided into two parts to be analyzed separately: a period inclusive of the 1963-1968 sewage diversion efforts and the 1976 establishment of *Daphnia* in the lake (1962-1977; LW$_1$) and a period following sewage diversion and *Daphnia* establishment (1976-1991; LW$_2$).

English Channel

The Western Channel Observatory, Plymouth, UK, has been monitoring the weekly abundance of English Channel (EC) zooplankton since 1988 and phytoplankton since 1992 at station L4, 18.5 km southwest of Plymouth. Eloire et al. (2010) and Southward et al. (2005) describe data collection methods for zooplankton and for phytoplankton, respectively. Time-points including both zooplankton and phytoplankton abundance data from 1992-2007 were used for analysis in this study.

Chesapeake Bay

The Chesapeake Bay Program oversees the monitoring of water quality and living resources in the Chesapeake Bay (CB) watershed. Data-series have been collected at numerous monitoring stations throughout the bay since 1984. We selected three stations representative of the lower (CB$_L$), middle (CB$_M$),

and upper (CB$_U$) portions of the estuary from among the stations representing the longest and most complete time-series.

Data preparation

Following the data preparation methods of previous studies that analyzed plankton data with MAR (e.g., Ives et al. 1999; Hampton et al. 2006; Scheef et al. 2012), the number of potential parameters to be included in the analysis was reduced by assigning the zooplankton and phytoplankton species in each dataset to general groups defined by similarities in taxonomy, feeding behavior, and susceptibility to predation. Rare plankton groups and those containing species unlikely to have been sampled adequately were excluded. The weekly plankton group abundances and driver values in each dataset were then averaged into monthly time-steps (Hampton et al. 2006, 2008), and each time-series was reduced to 170 consecutive time-steps to standardize the number of data-points included in the MAR analysis among all seven datasets.

Plankton groups and environmental variables in each dataset were categorized as either variates or covariates for inclusion in the MAR model (Ives et al. 2003). Variates are factors expected to affect their own dynamics and the dynamics of other variates (e.g., species abundances are typically treated as variates). Covariates may affect the dynamics of variates but are unlikely to be correspondingly influenced by them (e.g., temperature and salinity would be treated as covariates). With the exception of meroplankton, we considered each plankton group as a variate in the model, and all environmental parameters were considered covariates. Meroplankton groups were considered covariates because their abundance can be strongly influenced by benthic, rather than pelagic, processes. A summary of the variates and covariates selected for inclusion in MAR analysis for each dataset is given in Table 2.

The monthly data were log-transformed after replacing each plankton abundance value of zero with a random number between zero and half the lowest non-zero value for the group (Hampton et al. 2006). The logged data were then standardized to dimensionless units (Z-scores) such that model results could be directly compared between plankton groups. To remove the mean seasonal trends from the time-series, and thereby reduce the possibility of mistaking seasonal successions for interactions between groups, the Z-score value for each monthly time-step in a dataset was calculated by first subtracting the plank-

Table 2. Plankton groups and environmental drivers included in the MAR analysis for each time-series and their classification as variates or covariates in the model.

Time-series	Variates	Covariates
Lake Baikal		
	Copepod	Water temperature
	Rotifer	Arctic Oscillation
	Bosmina & Daphnia	El Niño Southern Oscillation
	Chrysophytes	Snow
	Cryptophytes	
	Cyanobacteria	
	Diatoms	
	Dinoflagellates	
	Green algae	
Lake Washington		
	Diatoms	Water temperature
	Green algae	Total phosphorus
	Oscillatoria	Neomysis
	Cyclops	
	Daphnia	
	Diaptomus	
	Non-daphnid cladocerans	
English Channel		
	Cnidarians	Meroplanktonic grazers
	Chaetognaths	Cirripedia
	Amphipods	Decapod larvae
	Small calanoid copepods	Water temperature
	Amphipods	
	Large calanoid copepods	
	Small calanoid copepods	
	Cyclopoid copepods	
	Poecilostomatoid copepods	
	Harpacticoid copepods	
	Diatoms	
	Dinoflagellates	
	Other algae	
Chesapeake Bay		
	Cnidarians[*]	Meroplanktonic grazers
	Chaetognaths[*]	Cirripedia
	Amphipods	Decapod larvae
	Small calanoid copepods	Water temperature
	Cladocerans	Salinity
	Cyclopoid copepods	Dissolved oxygen
	Harpacticoid copepods	
	Large calanoid copepods[*]	
	Poecilostomatoid copepods[**]	
	Ostracods[**]	
	Diatoms	
	Dinoflagellates	
	Other algae	

[*]Lower Chesapeake only
[**]Middle and upper Chesapeake only

ton group's corresponding multi-year mean abundance for the month, and then dividing the difference by the multi-year standard deviation for the month. Accounting for seasonal trends in the data also ensured that each time-series conformed to the assumption of process stationarity.

MAR application

The MAR framework (Ives et al. 2003) can be conceptualized as a series of linear regressions in which the abundance value of each plankton group at each time-step is a function of the values of all model variates and covariates at the previous time-step. The matrix notation of the formula is

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

where, for p interacting groups (variates) and q covariates (exogenous drivers), $\mathbf{X}_t$ is a $p \times 1$ vector of the groups' Z-scored abundance values at time t, $\mathbf{A}$ is a $p \times 1$ vector of intrinsic productivities (all equal to zero here because we are using Z-scored values), $\mathbf{B}$ is a $p \times p$ matrix of coefficients representing interactions among the variates, $\mathbf{X}_{t-1}$ is a $p \times 1$ vector of the Z-scored abundance values for each group at time $t - 1$, $\mathbf{C}$ is a $p \times q$ matrix of coefficients representing the effects of covariates on the variates, $\mathbf{U}_{t-1}$ is a $q \times 1$ vector of covariate values at time $t - 1$, and $\mathbf{E}$ is a $p \times 1$ vector of process errors with mean 0 and variance-covariance matrix $\mathbf{Q}$. The diagonal elements of $\mathbf{B}$ contain the density-dependent interaction terms for each variate (i.e., the effect of each variate on itself); the off-diagonal elements are the effects of the plankton groups on one another.

Following Ives et al. (1999), we used Akaike's Information Criterion (AIC) to evaluate the fit of a suite of potential models that could be constructed from each dataset. To find the "best-fit" model structure for each dataset, 100 models were randomly constructed by including or excluding coefficients ($\mathbf{B}$ and $\mathbf{C}$) with equal probability, and the best-of-100 model with the lowest AIC was retained. The process was repeated 100 times to generate 100 best-of-100 models from which the one with the lowest AIC was retained, resulting in a single model structure with the lowest AIC of 10,000 random models. Coefficients that were included in less than 15% of the best-of-100 models were dropped (Ives et al. 1999), and the entire 10,000 model selection process was repeated with the retained coefficients until all coefficients in the increasingly refined best-fit model fell above the 15% inclusion cut-off (i.e., none were dropped). The 95% confidence intervals for the coefficients in the best-fit model were obtained by bootstrapping ($n = 500$), and coefficients with confidence intervals that overlapped zero were eliminated (Hampton et al. 2006; Hampton and Schindler 2006). It is this final "bootstrapped" model from which inferences about community dynamics are drawn.

Assessment

MAR performance evaluation

We explored the behavior of MAR analysis and assessed its performance with each of our selected freshwater and marine

time-series by measuring the consistency of the best-fit and bootstrapped models generated among multiple runs of the analysis. Models were fit in Matlab using code made publicly available by Ives et al. (2003) through Ecological Archives, and the results were analyzed in R. MAR code is now also available for R in the MAR1 (Scheef 2013) and MARSS (Holmes et al. 2013) packages. MAR was applied to each 170 time-step dataset 10 times to generate 10 replicate best-fit and bootstrapped models to compare for consistency. We initiated the MAR models with a full suite of potentially interacting taxa, grouped as described above (Table 1). To simplify the comparisons of MAR performance across datasets, we did not restrict the interactions among the compartments as would normally be done (Ives et al. 2003), so biologically implausible interactions remained as possibilities in the model selection process.

There are two points in MAR analysis where the structure of the final, bootstrapped model for a particular dataset could readily diverge between replicates: 1) *different best-fit models can be retained* by the random selection process, and 2) *different coefficients can be removed* from the best-fit model by the bootstrap. Either of these factors or a combination of both may result in imperfect correlation among replicate MAR models for the same dataset.

Different best-fit models selected

Imperfect correlation among replicate best-fit MAR models indicates that the random model selection process ultimately generated different results despite the input of identical data. The selection of different models may result from the presence of weak interactions or otherwise low power to detect interactions due to possible overparameterization. Either of these circumstances would result in an increased likelihood of coefficients randomly falling above or below the 15 "hit" cut-off for inclusion during each iteration of the random model selection process (*see* "Materials and procedures"), which in turn could change the direction of the search or affect the point at which it ends.

Different coefficients eliminated by bootstrap

The model bootstrap is designed to eliminate coefficients that are not significantly different from zero; it is likely that coefficients representing weak or highly variable interactions will have bootstrapped confidence intervals that overlap zero and will thereby be removed by this process (Hampton et al. 2006). The bootstrap process may therefore improve consistency of the final model by buffering any variability in the random model selection process that results from weak interactions. At the same time, however, the bootstrap can be a source of model variability. When interaction coefficients with confidence intervals that nearly or barely overlap zero are present in the best-fit model, slight variations in their confidence limits—caused by variation in the best-fit coefficients carried through to the bootstrap or by random variation in the bootstrap process itself—could cause those interactions to blink in and out of replicate bootstrapped MAR models.

A summary of the MAR models generated for each freshwater and marine dataset is given in Table 3. The marine dataset MARs were initiated with more coefficients than the freshwater MARs, and similarly, the marine datasets contained more non-zero interaction coefficients in their best-fit models than the freshwater datasets. Overall, the final bootstrapped models included at least 35% fewer interactions than their corresponding best-fit models (Table 3). The proportions of interactions excluded by the bootstrap overlapped between the freshwater and marine datasets but were generally higher for the marine datasets. The marine CB_L and CB_U best-fit models were most notably altered by bootstrapping, with > 70% of their best-fit interactions ultimately excluded from their final bootstrapped models.

Replicate MAR model consistency was higher for the freshwater datasets than the marine datasets (Fig. 1). The mean τ-values for replicate best-fit models were above 0.98 for all three freshwater datasets and ranged from 0.83 to 0.93 for the marine datasets. Bootstrapping generally resulted in decreased consistency relative to the best-fit models, with the exception

Table 3. Total coefficients in the best-fit and bootstrapped MAR models for each dataset, mean number of nonzero interaction coefficients (*n* = 10) in their best-fit models and retained in their bootstrapped models, and the proportion of interaction coefficients removed through bootstrapping. Values for the marine dataset models in which copepod groups were combined into one group are shown in parentheses. Total coefficients include biologically implausible relationships that would have been excluded as potential interactions in studies focusing on food web dynamics (e.g., Hampton et al. 2006, 2008).

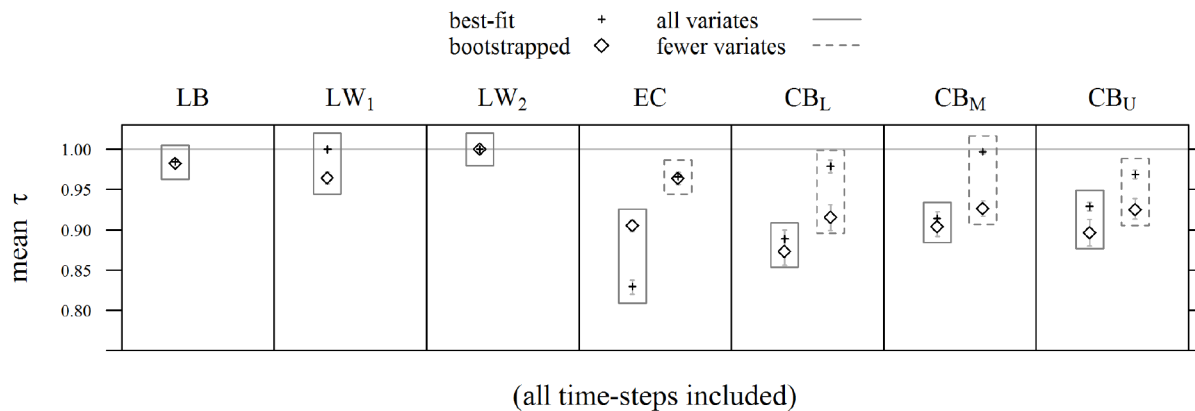| Dataset	Total coefficients	Nonzero coefficients best-fit	Nonzero coefficients bootstrapped	Proportion removed
LB	117	41.7	20.5	0.51
LW_1	70	27.0	16.9	0.37
LW_2	70	21.0	13.0	0.38
EC	252 (140)	79.8 (35.4)	40.2 (21.4)	0.50 (0.40)
CB_L	187 (112)	46.5 (34.1)	11.9 (9.1)	0.74 (0.73)
CB_M	160 (91)	51.8 (30.1)	23.3 (19.1)	0.55 (0.37)
CB_U	160 (91)	47.5 (30.6)	10.8 (9.5)	0.77 (0.69)

(all time-steps included)

Fig. 1. Mean -values (± 95% CI) for the best-fit and bootstrapped MAR models of each 170 time-step dataset and the condensed variate versions of the marine datasets. The horizontal gray line marks = 1.

of LW_2, where there was no change, and EC, where the τ-value increased by 0.08. However, all freshwater bootstrapped model τ-values remained above 0.96, whereas the marine bootstrapped model τ-values were all below 0.91.

The bootstrapped model consistencies indicate how much within-analysis variability (net from the best-fit model selection and bootstrap process) occurred for each dataset. Overall, less analysis variability occurred for the freshwater datasets than the marine datasets, which could suggest: 1) that the marine models were overparameterized and the number of variables included in the analysis should be reduced, 2) that the interactions among plankton in open marine systems are weaker or otherwise more difficult to detect, such that longer time-series are required for MAR analysis, or 3) that characteristics in addition to taxonomic and trophic similarities should be considered for grouping marine species.

MAR performance and number of variates

The marine datasets contain higher numbers of variates than the freshwater datasets, so we attempted to determine the degree to which this factor affected MAR performance. We repeated the sequence of MAR analyses described above on each marine dataset after combining the separate copepod groups into one (see model coefficient summary in Table 3), and compared the condensed model consistency results with the original model consistency results.

Reducing the number of variates included in the marine MAR analyses improved the consistency of the best-fit model results nearly to that of the freshwater datasets ($\tau > 0.96$; Fig. 1). However, the bootstrapped model consistencies for these condensed variate marine datasets did not notably improve relative to the bootstrapped model consistencies of the original marine datasets (τ-value increase ranged from 0.02–0.04), with the exception of EC (τ-value increase of 0.06).

Improvement of model consistency after the number of variates included in the MAR analysis for each marine dataset was reduced suggests that overparameterization is partly responsible for within-analysis variability in the marine

datasets. The EC dataset responded particularly strongly to a reduction in parameters, with consistency improving to levels approaching the freshwater models. Although the MAR performance with the three CB datasets also improved somewhat with the reduction of variates, bootstrapped model consistencies remained lower than those of EC and the freshwater datasets. The removal of high proportions of interactions from their best-fit models by the bootstrap process (Table 3) suggests weak interactions may have contributed to the poor model correlations for those datasets. Whereas lengthening these time-series so MAR can better detect weak interactions may help improve model consistency, it also seems likely that modifying the configuration of the model variables could help achieve more informative results. For example, the salinity tolerances of species in estuarine systems such as CB may need to be considered when grouping taxa for MAR analysis.

MAR performance and number of time-steps

To explore how the dependence of MAR performance on time-series length differs between datasets, we examined the degradation of replicate MAR model consistency as the number of included time-steps in each dataset was reduced. We executed jackknife MAR analyses in which data-points (aligned t and $t-1$ values) were randomly removed from the complete time-series at the beginning of each replicate run ($n = 10$). The length of the degraded datasets ranged from 100% to 60% (in 5% increments) of the original 170-step time-series. Model consistency within each set of 10 replicates was calculated as the mean Kendall's rank correlation (τ-value) of all 45 possible replicate model comparisons within the set.

As increasing proportions of data-points were dropped from the time-series during the jackknife analyses, consistency of MAR models (both best-fit and bootstrapped) decreased for all datasets (Fig. 2). Mean τ-values for the marine results were lower overall and decreased to dramatically lower levels as data-points were removed than for the freshwater results. Mean τ-values for replicate marine models ultimately fell below 0.55 at the 60% jackknife interval, whereas τ-values
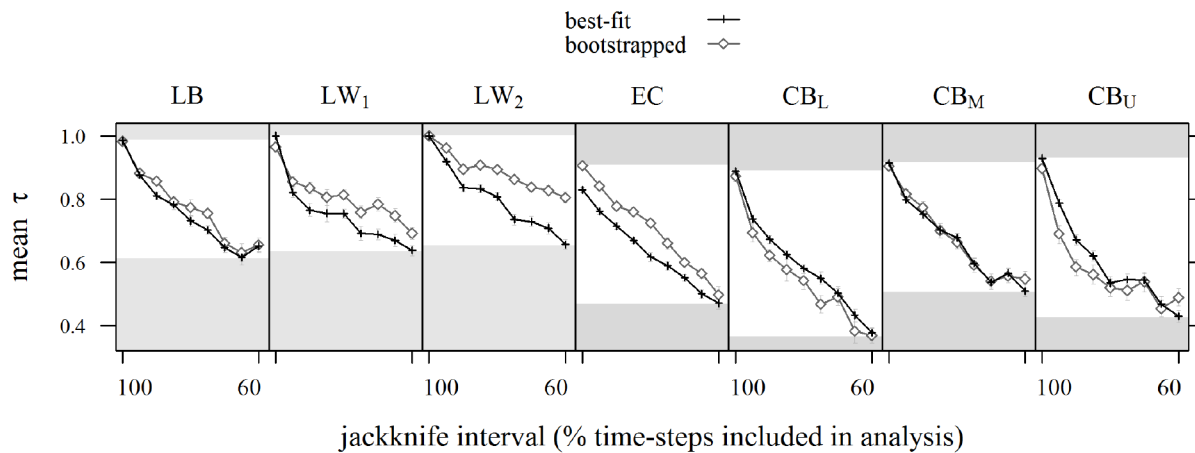
Fig. 2. Mean -values (± 95% CI) at each jackknife interval for the best-fit and bootstrapped MAR models of each dataset. Shaded areas emphasize maximum and minimum values.
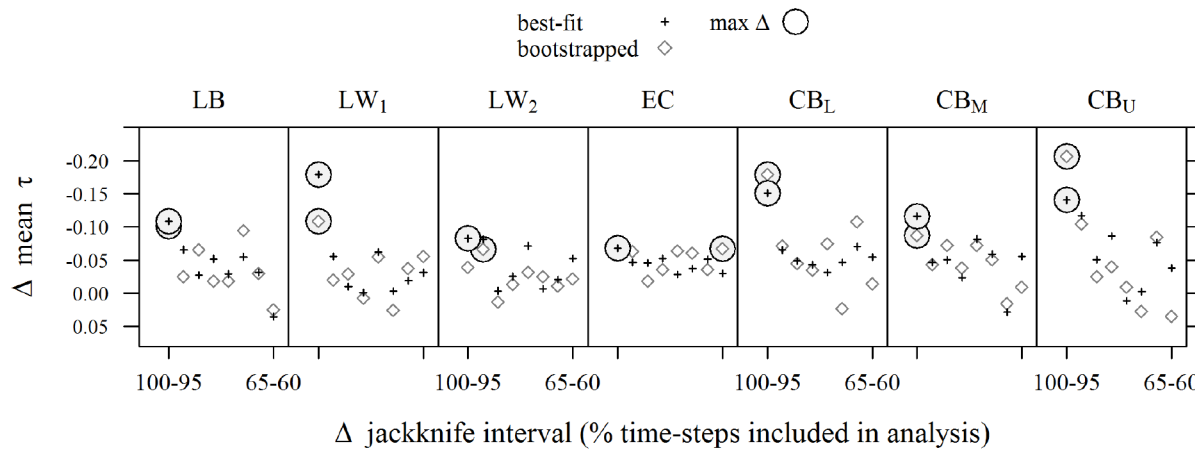

Fig. 3. Change in mean τ between each 5% jackknife interval for the best-fit and bootstrapped models of each dataset. The largest changes for each dataset are circled. The 95% confidence intervals are all < 0.015 and generally do not overlap unless symbols overlap.

remained above 0.6 for all freshwater model sets. The magnitude of these decreases did not significantly differ between the best-fit models and boot-strapped models (–0.91 versus –0.80; ANCOVA, p = 0.48) but was higher for the marine results than for the freshwater results (–0.98 versus –0.69; ANCOVA, p < 0.001). The more dramatic declines in marine model consistencies relative to freshwater indicate that time-series length was more critical to the generation of reliable MAR models for the marine datasets, potentially due to weaker interactions among the marine plankton groups.

Generally, for all the datasets, declines in result consistency with the removal of data-points were most dramatic between the 100% and 95% time-step inclusion increments, followed by steady, more gradual decreases between the other increments (Fig. 3). Although dataset length is a factor of potential influence on MAR performance at all jackknife intervals from 100% to 60%, variability in which data-points were included in each replicate analysis could have affected MAR result consistency only when time-steps were removed in the 95% to

60% inclusion intervals. The change in MAR model consistency between 100% and 95% of time-points included, therefore, reflects the continuing influence of dataset length on MAR performance and the introduction of influence from randomly including or excluding particular values in the dataset (Fig. 3). More dramatic changes in model consistency between the 100% and 95% data inclusion intervals, relative to changes between the other intervals, represent stronger effects of a few points in the dataset on MAR results (possible sporadic interactions between groups or intermittent driver influences). Less dramatic changes in bootstrapped model consistency relative to best-fit model consistency indicate that bootstrapping successfully helped to exclude interactions that were weakly supported by a few data-points. More dramatic changes in bootstrapped relative to best-fit model consistency suggest an aggravation of the within-analysis variability discussed above.

Marine datasets appeared no more likely to contain influential points than freshwater datasets. LW_1, CB_L, and CB_U MAR

results appeared to be the most influenced by varying the data values analyzed (Fig. 3). LW_1, however, was the only dataset of these three where bootstrapping served its purpose of buffering variability in the final model caused by influential time-steps, suggesting that MAR analysis was more robust with that dataset. The datasets least influenced by specific time-step values were LW_2 and EC. Therefore, although result consistency for EC was generally low (Fig. 2), the inclusion of influential points did not appear to be a primary factor reducing the appropriateness of MAR for that dataset.

MAR performance summary

MAR analysis performed relatively well with the three freshwater datasets we tested, which is perhaps surprising given that we used parameter sets that had been designed for use with much longer time-series from those lakes, and that we did not apply the restrictions imposed on the analysis in previous studies (Hampton et al. 2006, 2008). Although the LB dataset appeared to include some influential points, result inconsistency due to within-analysis variability was relatively low and bootstrapped models generated from that dataset were nearly identical. Interestingly, MAR behavior differed between the two Lake Washington datasets. Within-analysis variability was higher with LW_1, causing its bootstrapped models to vary slightly, while the LW_2 models were identical between replicates (Fig. 1). LW_1 also appeared to include time-steps that had strong influences on its MAR results, suggesting this dataset may have been on the threshold of its minimum ideal length for MAR analysis. The establishment of *Daphnia* as a strong interactor in the LW_2 portion of the Lake Washington time-series (Table 1) may have contributed to better MAR performance with that dataset relative to LW_1.

The marine datasets appeared to present challenges to MAR analysis. For the EC dataset, variability within the analysis was high. The presence of consistent, but weak interactions may have been partially responsible for the poor MAR performance with EC, but the majority of variability seemed to be attributable to overparameterization. Overparameterization was also apparent for the three CB datasets, but their MAR result consistency remained relatively low after the reduction of variables. Weak and intermittent interactions within this estuarine system, possibly due to unsuitable species groupings, appear to have contributed to poor MAR performance with the CB datasets. Grouping taxa exclusively by taxonomic and trophic identity, as has been done in freshwater MAR studies, may not provide the most suitable parameterizations for marine datasets in which associations with environmental drivers are less aligned with taxonomic or trophic delineations.

Comments and recommendations

As anticipated, MAR results showed greater consistency for freshwater datasets relative to marine datasets, suggesting that standard MAR may be better-suited to the type of data that are available from freshwater plankton research. However, the marine results suggest that some of the factors affecting MAR performance may be under the investigator's control in the analysis, rather than simply being a reflection of more dynamic environmental conditions. Specifically, marine MAR model consistency improved after reducing the number of potential interactors included in the analysis, and MAR performance varied not only between freshwater and marine datasets, but among datasets of each of those data types.

The poor consistency of marine MAR results appeared to result from overparameterization, the presence of weak interactions, or both. The risk of overparameterization can be reduced by including fewer variables in the model, and weak interactions may become more apparent if plankton groups are defined differently. It therefore appears that an investigator focused on a specific ecosystem and question could dramatically improve MAR performance with their dataset by adjusting how species are assigned into groups to parameterize the model.

The consolidation of species in a dataset into as few groups as possible is important to help avoid overparameterization (Ives et al. 1999; Hampton and Schindler 2006). However, this can be quite difficult for marine plankton datasets given the great taxonomic and functional diversity of the species they may include. For example, in the freshwater datasets analyzed here, copepod species were combined into either one or 2 groups based on similarities in taxonomy, feeding behavior, and susceptibility to predation (Hampton et al. 2006, 2008). The use of these same criteria to group the zooplankton species in the marine datasets resulted in five distinct copepod groups for inclusion in the MAR models (Table 2).

Although we lumped the copepod groups together into one variate to test whether reducing the number of parameters would improve the performance of the analysis, the models suggest that this would not be appropriate for a meaningful assessment of plankton dynamics. According to the MAR models produced, it appears that the various marine copepod groups do, in fact, interact differently with the plankton community and that information is lost by combining them into one. For example, the bootstrapped model of the EC dataset with the five separate copepod groups indicates that some types of copepods negatively affect the abundance of diatoms and dinoflagellates, but the model with the single combined copepod group suggests there were no effects of copepods on primary producers.

How, then, should species grouping be done to result in the fewest number of parameters (leading to more stable MAR results) and a minimal loss of information? Some of these decisions will be system- and question-specific and will rely heavily on the investigator's knowledge of the ecosystem. In some systems, there may be environmental drivers that can inform taxonomic groupings. For example, in an estuarine system such as the Chesapeake Bay, it may be useful to take the salinity tolerances of specific species into consideration during grouping. In a more open ocean system, one may want

to consider the possible water mass origins of some species (nearshore or offshore, northern or southern, etc.; Francis et al. 2012). If interactions at higher or lower trophic levels are of primary interest, coarser lumping of primary producers or predators, respectively, may be appropriate. Analyses that search for underlying trends in multivariate data-series, such as dynamic factor analysis, may also provide useful clues to how plankton species could be grouped. The results of a preliminary dynamic factor analysis of the EC plankton groups suggest, however, that these analyses should be used with caution as they can potentially lump strong interactors from different trophic levels together (E.E. Holmes and L.P. Scheef unpubl. data).

Conclusions

The exploration of MARs in freshwater systems has provided a blueprint for analyzing marine datasets with MARs, and our results suggest some specific aspects of the process that will need to be altered for marine plankton datasets in particular, and considered more generally for datasets beyond plankton. Specifically, approaches to parameterizing freshwater MARs lead to overparamaterization in marine MARs; we suggest that investigators move beyond taxonomic and trophic groupings to consider coarse categorizations and environmental tolerances as appropriate for each particular system and question of focus. On the horizon, analysis of ecological data with modified MAR frameworks that increase its flexibility—such as recent applications of state-space versions of the MAR model (Lindegren et al. 2009; Ward et al. 2010)—may help to improve its performance with noisy or incomplete time-series and facilitate the detection of species interactions and environmental drivers (Hampton et al. in press). As monitoring programs continue to be initiated and mature throughout the world, an increased understanding of the existing limitations, and potential refinements of analytical frameworks such as MAR will maximize what we are able to learn from these hard-won data.

References

Bascompte, J., C. J. Melián, and E. Sala. 2005. Interaction strength combinations and the overfishing of a marine food web. Proc. Natl. Acad. Sci. USA 102:5443-5447 [doi:10.1073/pnas.0501562102].

Cabana, G., and J. B. Rasmussen. 1994. Modelling food chain structure and contaminant bioaccumulation using stable nitrogen isotopes. Nature 372:255-257 [doi:10.1038/372255a0].

Edmondson, W. T., and J. T. Lehman. 1981. The effect of changes in the nutrient income on the condition of Lake Washington. Limnol. Oceanogr. 26:1-29 [doi:10.4319/lo.1981.26.1.0001].

———, and A. H. Litt. 1982. Daphnia in Lake Washington. Limnol. Oceanogr. 27:272-293 [doi:10.4319/lo.1982.27.2.0272].

———, S. E. B. Abella, and J. T. Lehman. 2003. Phytoplankton in Lake Washington: long-term changes 1950-1999. Arch. Hydrobiol. Suppl. 139:275-326.

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. J. Plankton Res. 32:657-679 [doi:10.1093/plankt/fbq009].

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 Biol. 18:2498-2508 [doi:10.1111/j.1365-2486.2012.02702.x].

Hampton, S. E., and D. E. Schindler. 2006. Empirical evaluation of observation scale effects in community time series. Oikos 113:424-439 [doi:10.1111/j.2006.0030-1299.14643.x].

———, M. D. Scheuerell, and D. E. Schindler. 2006. Coalescence in the Lake Washington story: Interaction strengths in a planktonic food web. Limnol. Oceanogr. 51:2042-2051 [doi:10.4319/lo.2006.51.5.2042].

———, 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. Glob. Change Biol. 14:1947-1958 [doi:10.1111/j.1365-2486.2008.01616.x].

———, E. E. Holmes, L. P. Scheef, M. D. Scheuerell, S. L. Katz, D. E. Pendleton, and E. J. Ward. In press. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. Ecology.

Holmes, E., E. Ward, and K. Wills. 2013. MARSS: Multivariate autoregressive state-space modeling. R package v. 3.4. <http://cran.r-project.org/web/packages/MARSS>.

Huber, V., and U. Gaedke. 2006. The role of predation for seasonal variability patterns among phytoplankton and ciliates. Oikos 114:265-276 [doi:10.1111/j.2006.0030-1299.14753.x].

Ings, T. C., and others. 2009. Ecological networks—beyond food webs. J. Anim. Ecol. 78:253-269 [doi:10.1111/j.1365-2656.2008.01460.x].

Ives, A. R. 1995. Predicting the response of populations to environmental change. Ecology 76:926-941 [doi:10.2307/1939357].

———, S. R. Carpenter, and B. Dennis. 1999. Community interaction webs and zooplankton responses to planktivory manipulations. Ecology 80:1405-1421 [doi:10.1890/0012-9658(1997)080[1405:CIWAZR]2.0.CO;2].

———, B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. Ecol. Monogr. 73:301-330 [doi:10.1890/0012-9615(2003)073[0301:ECSAEI]2.0.CO;2].

Izmest'eva, L. R. 2006. Lake Baikal plankton. KNB data registry. <http://knb.ecoinformatics.org/knb/metacat/nceas.290.3/nceas>.

Kozhova, O. M., and L. R. Izmest'eva. 1998. Lake Baikal: evo-

lution and biodiversity. Backhuys Publishers.

Layman, C. A., D. A. Arrington, C. G. Monta-a, and D. M. Post. 2007. Can stable isotope ratios provide for community-wide measures of trophic structure? Ecology 88:42-48 [doi:10.1890/0012-9658(2007)88[42:CSIRPF]2.0.CO;2].

Lindegren, M., C. Möllmann, A. Nielsen, and N. C. Stenseth. 2009. Preventing the collapse of the Baltic cod stock through an ecosystem-based management approach. Proc. Natl. Acad. Sci. 106:14722-14727 [doi:10.1073/pnas.0906 620106].

Pace, M. L., J. J. Cole, S. R. Carpenter, and J. F. Kitchell. 1999. Trophic cascades revealed in diverse ecosystems. Trends Ecol. Evol. 14:483-488 [doi:10.1016/S0169-5347(99)01723-1].

Paine, R. T. 1980. Food webs: linkage, interaction strength and community infrastructure. J. Anim. Ecol. 49:667-685 [doi:10.2307/4220].

Scheef, L. P. 2013. MAR1: Multivariate autoregressive modeling for analysis of community time-series data. R package ver. 1.0. <http://cran.r-project.org/web/packages/MAR1>.

———, 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. Limnol. Oceanogr. Methods 10:54-64 [doi:10.4319/lom.2012.10.54].

Soulé, M. E., J. A. Estes, B. Miller, and D. L. Honnold. 2005. Strongly interacting species: conservation policy, management, and ethics. BioScience 55:168.

Southward, A. J., and others. 2005. Long-term oceanographic and ecological research in the western English Channel. Adv. Mar. Biol. 47:1-105 [doi:10.1016/S0065-2881(04) 47001-1].

Ward, E. J., H. Chirakkal, M. González-Suárez, D. Aurioles-Gamboa, E. E. Holmes, and L. Gerber. 2010. Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. J. Appl. Ecol. 47:47-56 [doi:10.1111/j.1365-2664.2009.01745.x].