Appendix D: Simulation studies of small-sample properties of MAR estimators

Simulation study 1: Sampling distributions of MAR model parameters and stability metrics under different magnitudes of trend

We conducted a small simulation study to investigate whether the trend covariate in eq. (1) impacted the estimation of several of the stability metrics calculated in the cover analysis. For each of the three habitats, we simulated data sets using the estimated values of **a**, **B**, and **C** in eq. (1) as the generative model. For the RS habitat, we used the average estimate of **a** across all sites. We simulated environmental variation and residual variation by sampling with replacement from the observed environmental vectors and the estimated residual vectors, respectively. Because our simulation focused on the effect of time, we ran simulations where **z** in the generative model equaled $k\mathbf{z}$, where k is a multiplier that diminished or amplified the trend by a factor of k=0, 0.5, 1, 1.5 or 2, and **z** equaled its estimated value. We refer to the k as the "trend multiplier". Initial values for the cover composition were drawn from the estimated quasi-stationary distribution for the first year of our study. Thus, we had 15 total simulation scenarios (three habitats crossed with five values of k). We simulated 1000 data sets for each simulation scenario, with each data set lasting for 21 time steps (the same duration as the actual data). Rare simulations that generated an estimated **B** matrix with a spectral radius greater than 1 were discarded. We report the mean, interquartile range, and 10^{th} and 90^{th} percentiles of the empirical sampling distribution for several model parameters and derived stability metrics.

Figure D1 shows the empirical sampling distribution for the elements of **a**, **B**, **C**, and **z** for each habitat and each value of *k*, along with actual values from the generative models. Not surprisingly, estimators are biased. This bias is not surprising because it is known that the conditional least-squares estimators of autoregressive models are only asymptotically unbiased, and will be biased for short time series. The key feature of figure D1, however, is that the marginal sampling distribution of the elements of **a**, **B**, and **C** do not appear to depend on the value of *k*. The standard error of the elements of **z** increases as *k* increases, but any bias in the elements of **z** appears to be small.

Figure D2 shows the empirical sampling distribution of several derived metrics for the same simulation scenarios. Results suggest varying degrees of bias in derived quantities, although the magnitude of the bias only depends minimally on the strength of the trend. Means of the quasi-stationary distribution — either on the ilr-transformed scale or on the proportion scale — show negligible bias. The CV of proportional coral cover at the quasi-stationary distribution is slightly downwardly biased for the Tektite scenario, and slightly upwardly biased for the RS. Sensitivities of mean coral cover (again, on the proportion scale) to both environmental covariates and the trend all seem to be estimated with little or negligible bias.

The spectral radius of the **B** matrix is biased for all simulation scenarios for Tektite and Yawzi Point, and in these cases is negatively biased (that is, the true spectral radius is larger than the average estimated spectral radius). That the spectral radius is estimated with bias is perhaps not surprising, given the strong non-linearity inherent in calculating eigenvalues.

Nevertheless, the bias makes it clear that differences between spectral radii across habitats (fig. 2b of the main text) should be interpreted cautiously.

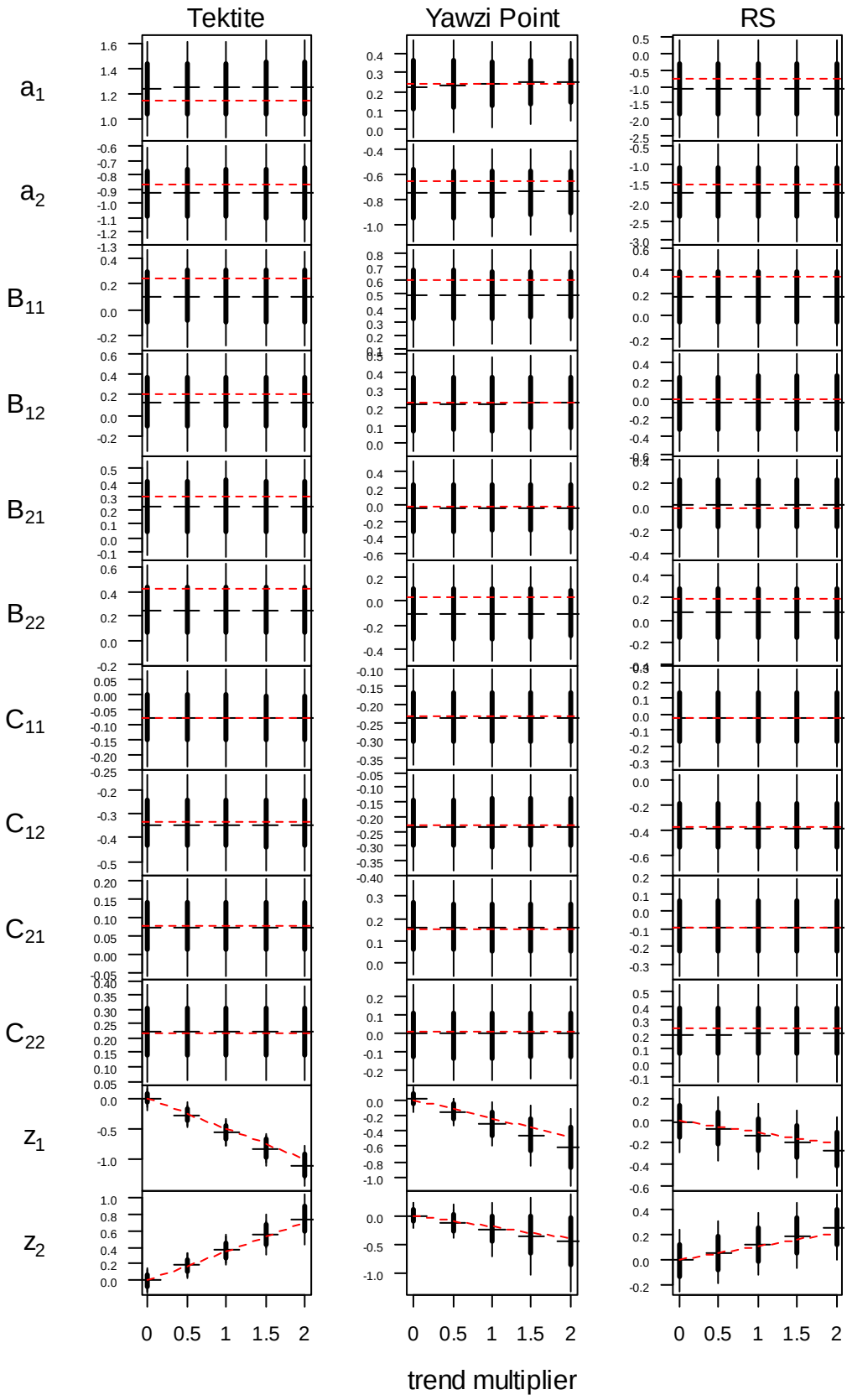
Tektite
Yawzi Point
RS
a1
a2
B11
B12
B21
B22
C11
C12
C21
C22
z1
z2
0
0.5
1
1.5
2
trend multiplier

*Figure D1 (previous page). Sampling distributions of the elements of **a**, **B**, **C**, and **z** for different simulated scenarios. Columns of panels correspond to the habitat that was used as the generative model, rows of panels show different parameters, and segments within panels show different values of the trend multiplier. Horizontal hashes show the average parameter estimate, thick vertical line segments span the interquartile range of the sampling distribution, and thin vertical line segments range from the 10^{th} percentile to the 90^{th} percentile of the sampling distribution. Red lines connect actual parameter values from the generative model.*

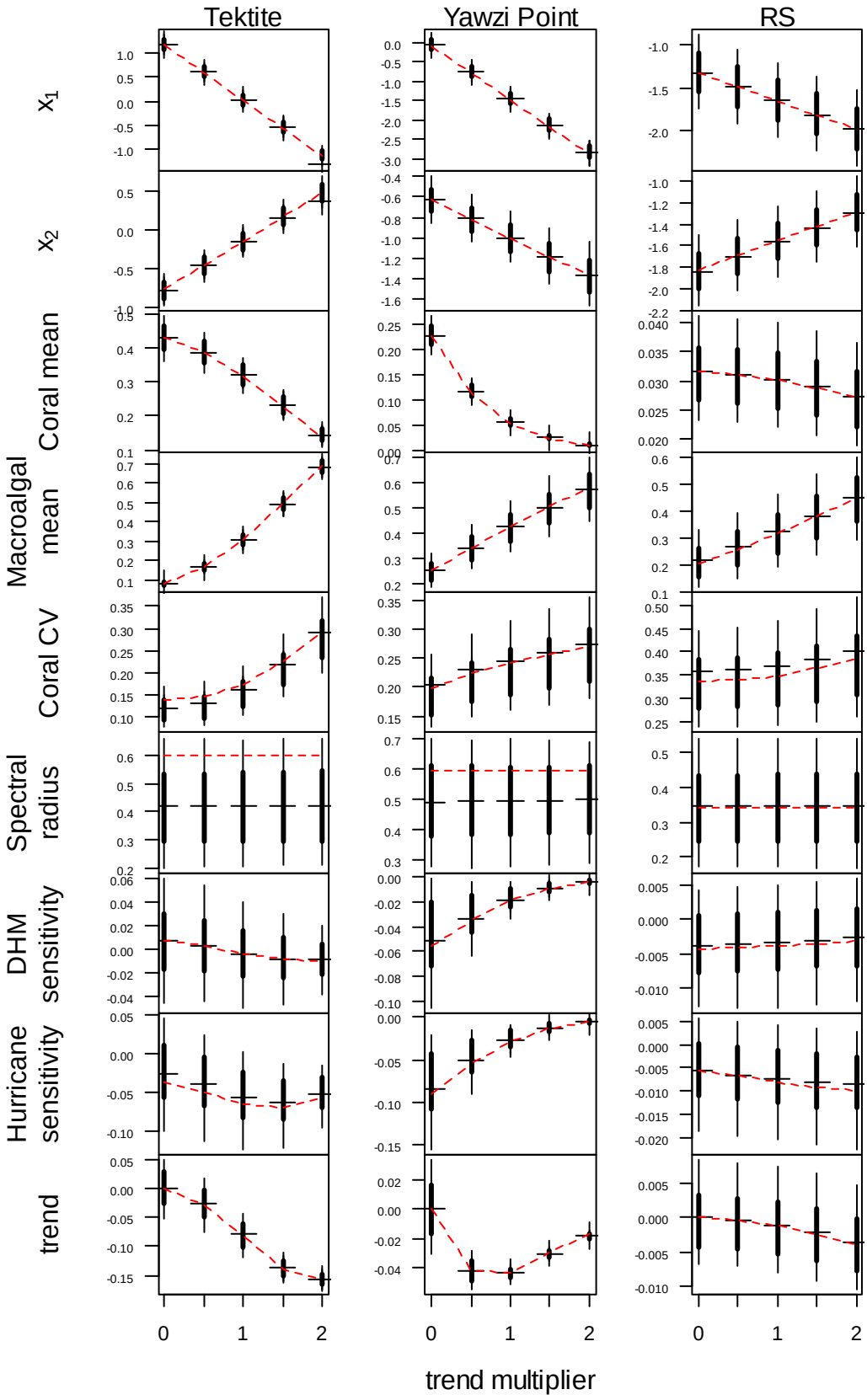
Tektite
Yawzi Point
RS
x_1
x_2
Coral mean
Macroalgal mean
Coral CV
Spectral radius
DHM sensitivity
Hurricane sensitivity
trend
trend multiplier

Figure D2 (previous page). Sampling distributions of derived quantities of interest for different simulated scenarios. Basic arrangement of panels is the same as Figure D1. First two rows: elements of $\boldsymbol{\mu}_x$ *(quasi-stationary distribution on the ilr-transformed scale). Third and fourth rows: elements of* $\boldsymbol{\mu}_p$ *, mean coral and macroalgal cover at the quasi-stationary distribution on the proportion scale. Fifth row: CV of coral cover at the quasi-stationary distribution, on the proportion scale. Sixth row: spectral radius. Seventh and eight rows: Sensitivity of average coral cover to environmental covariates, on the proportion scale (*$d\boldsymbol{\mu}_p / d\boldsymbol{\mu}_u \times \left(1/\boldsymbol{\mu}_p\right)$*). Ninth row: Trend of average coral cover per year, on the proportion scale (*$d\boldsymbol{\mu}_p / dt^* \times \left(1/\boldsymbol{\mu}_p\right)$*).*

Simulation study 2: Quality of approximate probability contours in figure 2 when $\boldsymbol{u}_t$ *and* $\boldsymbol{e}_t$ *are not normally distributed.*

Neither the MAR model (eq. 1) nor any of our results (eqq. 2 – 5) require a normality assumption for either the environmental covariates in $\mathbf{u}_t$ or the random errors in $\mathbf{e}_t$. However, the approximate probability contours for the quasi-stationary distribution shown in Fig. 2 are based on the assumption that the quasi-stationary distribution is multivariate normal on the ilr-transformed scale, which in turn relies on a normality assumption for both $\mathbf{u}_t$ and $\mathbf{e}_t$. To assess the accuracy of the approximate probability contours when $\mathbf{u}_t$ and $\mathbf{e}_t$ are not normally distributed, we simulated 5000 years of dynamics for the Tektite and Yawzi Point habitat from eq. (1), fixing the trend covariate at its 2012 value, and drawing $\mathbf{u}_t$ from its observed distribution and independently drawing $\mathbf{e}_t$ from the estimated residuals from each model.

Simulations were initiated from the estimated metric center of each quasi-stationary distribution, and the first 100 years of dynamics were discarded as a burn-in. Both $\mathbf{u}_t$ and $\mathbf{e}_t$ were sampled as vectors, thus preserving correlations between environmental covariates, and between the residuals in $\mathbf{e}_t$. These simulated dynamics provide a visualization of the exact quasi-stationary distribution for these two habitats, assuming that the actual distributions of $\mathbf{u}_t$ and $\mathbf{e}_t$ are identical to their empirical distributions (Fig. D3). The proportion of simulated data points that fall within the 50%, 80% and 95% probability contours shown in Fig. 1 of the main text are: 55.9%, 82.5%, and 95.2%, respectively, for Tektite; 57.3%, 87.2% and 94.0%, respectively, for Yawzi Point; and 56.4%, 81.5% and 93.2%, respectively, for the RS . Figure D4 shows a quantile-quantile plot of the Mahalanobis distances for the simulated dynamics on the ilr-transformed scale vs. theoretical quantiles from a χ^2_2 distribution. (This plot is the multivariate analog of the familiar normal probability plot used in residual analysis.)

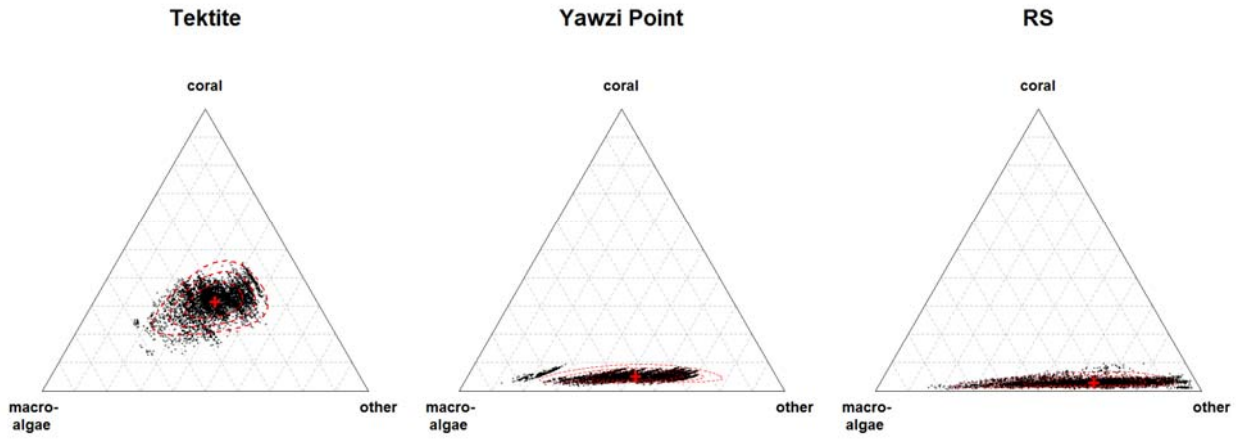

Figure D3. Triangle plots of 5000 years of simulated data from each of the three habitats, using the 2012 value of the trend covariate and the empirical distributions of $\mathbf{u}_t$ and $\mathbf{e}_t$. Red

dashed lines show approximate 50%, 80% and 95% probability contours for comparison, and are identical to those shown in Figure 1 of the main text.

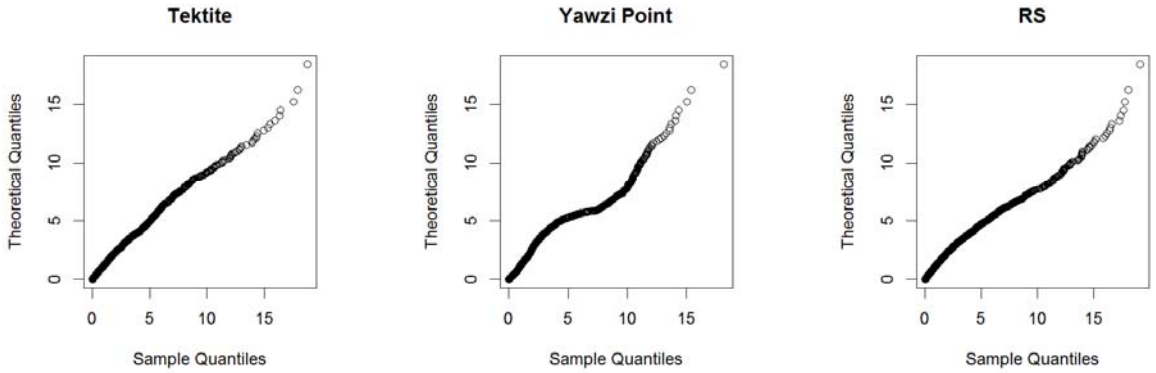

Figure D4. Quantile-quantile plots of the Mahalanobis distances of the simulated compositions (on the ilr-transformed scale) from the metric center of the quasi-stationary distribution for all three habitats, using a χ_2^2 *distribution for comparison. Departures from linearity suggest differences between the quasi-stationary distributions generated using empirical distributions of* $\boldsymbol{u}_t$ *and* $\boldsymbol{e}_t$ *(on the ilr-transformed scale) and their multivariate normal approximations.*

Taken together, these plots suggest that the approximate probability contours shown in Fig. 2 of the main text are reasonable approximations. There is some multimodality apparent in the distribution of simulated data at the Yawzi Point habitat, but this is likely a consequence of an anomalous residual (see Figure C1 of appendix C). This and other fine structure apparent in the empirical distribution of residuals is likely a consequence of the coarseness in the empirical distributions of $\mathbf{u}_t$ and $\mathbf{e}_t$ that arises from having a limited number of data points. It is unlikely that this coarseness would persist if more data were available. Thus, it seems appropriate to

view the normal-based probability contours as an approximation that captures the main features of the quasi-stationary distribution without over-fitting to idiosyncratic fine structure.