Stochastic dynamics of a warmer Great Barrier Reef

Jennifer K Cooper, Matthew Spencer, John F Bruno

Appendices (additional methods and results)

A.1 Data sources

Reef composition data come from the intensive surveys of the Australian Institute of Marine Science Long-Term Monitoring Programme. These surveys were carried out using video transects. There are 3 sites per reef, and 5 permanently-marked 50 m transects per site (Abdo et al. 2004, p. 5). On each transect, the benthic organisms present at 200 points are identified (Abdo et al. 2004, p.30).

Sea surface temperature (SST) and anomaly data were obtained from the CorTAD database Version 3 (http://www.nodc.noaa.gov/sog/cortad/, accessed 02/12/2011) (Selig et al. 2010). Weekly SST data at 4 km resolution were extracted using Matlab R2012a (The Mathworks, Inc., Natick, MA), and a mean SST and mean SST anomaly calculated for each calendar year. We obtained site-specific climatology data (Selig et al. 2010) by subtracting mean SST anomaly from mean SST for each site in each year.

We used centred and scaled (to zero mean and unit standard deviation) lagged SST anomaly and climatology (for the calendar year preceding the explanatory observation, and thus two years before the response observation) as explanatory variables in our models. We denote by $z_1(t) = \frac{v_1(t)-\bar{v}_1}{s_1}$ the centred and scaled lagged annual mean SST anomaly at time t, where $v_1(t)$ is the value of lagged annual mean SST anomaly at time t, $\bar{v}_1$ is the sample mean lagged annual mean SST anomaly (over all times and

locations), and s_1 is the sample standard deviation of lagged annual mean SST anomaly (over all times and locations). Similarly, we denote by $z_2(t) = \frac{v_2(t)-\bar{v}_2}{s_2}$ the centred and scaled lagged climatology at time t.

We also investigated other combinations of SST variables (lagged and unlagged SST, lagged climatology and anomaly separately, and unlagged climatology and anomaly, together or separately). The combination of lagged climatology and lagged anomaly outperformed all other SST variables, as measured by Akaike's Information Criterion (Table A1).

We downloaded spatial data on integrated local threats to reefs from http://www.wri.org/publication/reefs-at-risk-revisited#datasets (accessed 06/10/2012). Local threat categories were decided by expert judgement (Burke et al. 2011, p. 16). Most of the underlying data used to produce these categories were resolved to a 1 x 1 km grid, according to the metadata at http://www.wri.org/sites/default/files/reefs_at_risk_revisited_metadata_local_threats.xlsx and the technical notes at http://www.wri.org/sites/default/files/docs/reefs_at_risk_revisited_technical_notes.pdf (both accessed 2 June 2014). The exception, population growth was on a smoothed 3 x 3 km grid (Reefs at Risk Revisited technical notes, page 4). The effects of point sources of local threat were weighted by distance (Burke et al. 2011, p. 16; Reefs at Risk Revisited technical notes, Tables 2 and 4 and pages 7, 8, 10, 12).

These data were in vector format. We used the R packages rgdal (version 0.7-19) and rgeos (version 0.2-7) to obtain the local threat level for the polygon containing each reef. 39 pairs of observations came from reefs that were not in any polygon with a defined local threat level. All were less than 0.5 km from the closest polygon with a local threat level, so we assigned them the local threat level of this closest polygon (270 pairs of observations had low local threat level, 80 medium, 5 high (all from a single reef), and 9 very high (all from a single reef)). Locations of reefs and their local threat levels are shown in Fig. A1.

A.2 Short-term changes in reef composition

If the reef compositions $\mathbf{y}(t)$ and $\mathbf{y}(t+1)$ at two successive time points were vectors in an n-dimensional real space, we could simply subtract $\mathbf{y}(t)$ from $\mathbf{y}(t+1)$ to obtain a measure of change. However, $\mathbf{y}(t)$ and $\mathbf{y}(t+1)$ are vectors in an $(n-1)$-dimensional simplex, and simple subtraction may result in a vector that is not a composition. Instead, we use the perturbing vector $\mathbf{p}(t) = \mathcal{C}\left(\frac{y_1(t+1)}{y_1(t)}, \frac{y_2(t+1)}{y_2(t)}, \frac{y_3(t+1)}{y_3(t)}\right)$, where $\mathcal{C}(\mathbf{w}) = \frac{\mathbf{w}}{\sum_i w_i}$ denotes the closure operation for a non-negative vector $\mathbf{w}$ (Aitchison 1986, p. 42).

A.3 Transformation of compositions

Denote by $\mathbf{x}(t) = \text{ilr}\,\boldsymbol{y}(\text{t})$ the isometric log-ratio transformed composition at time t. The elements of $\mathbf{x}(t)$ are

$$x_1(t) = -\frac{1}{\sqrt{2}}\log\frac{y_1(t)}{y_2(t)}, \qquad \text{(A.1)}$$

which increases as the ratio of algae to coral increases, and

$$x_2(t) = -\frac{2}{\sqrt{6}} \log \frac{\sqrt{y_1(t) y_2(t)}}{y_3(t)}, \quad \text{(A.2)}$$

which increases as the ratio of 'other' to the geometric mean of coral and algae increases (throughout, log denotes natural logarithm). The origin of this coordinate system corresponds to equal proportions of coral, algae, and 'other'. In addition, $\text{ilr}\, \mathbf{p}(t) = \text{ilr}\, \mathbf{y}(t+1) - \text{ilr}\, \mathbf{y}(t)$.

We used the R package compositions version 1.20-1 (van den Boogaart and Tolosana-Delgado 2008) to carry out these transformations.

A.4 Model details

Our model is

$$\text{ilr}\, \mathbf{p}(t) = \mathbf{c} + \mathbf{A}\mathbf{x}(t) + \boldsymbol{\beta}_1 z_1(t) + \boldsymbol{\epsilon}(t). \quad \text{(1)}$$

The first term on the right of Equation 1 is

$$\boldsymbol{c} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_2 z_2 + \boldsymbol{\Gamma}\mathbf{r}, \quad \text{(A.3)}$$

which is constant over time and consists of three terms. The first is an intercept vector $\boldsymbol{\beta}_0$. The second term is the effect of centred and scaled climatology z_2, with coefficient $\boldsymbol{\beta}_2$. The third term $\boldsymbol{\Gamma}\mathbf{r}$ is the effect of the categorical local threat levels. Local threat levels are represented by a 3×1 vector $\boldsymbol{r}$ of binary variables (assumed constant over time), only one of which was nonzero for a given reef, indicating whether the reef was in the Medium, High, or Very High category. The 2×3 matrix $\boldsymbol{\Gamma}$, with columns $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3$, is the effect on the expected transformed perturbing vector of being in each of the local threat levels Medium, High, or Very High, relative to the Low threat level. Using Low as the reference level is

appropriate because for a reef to have Low local threat, it must receive the minimum possible score for every local stressor (Burke et al. 2011, p. 18), which may make Low qualitatively different from the 'other' categories. The vector $\mathbf{c}$ can be thought of as describing the deterministic aspect of reef dynamics in the absence of effects of reef composition.

The second term in Equation 1, $\mathbf{A}\mathbf{x}(t)$, describes the effects of current reef composition on short-term changes. The vector $\mathbf{x}(t)$ is the ilr-transformed reef composition at time t. The first column $\mathbf{a}_1$of the 2×2 matrix $\mathbf{A}$ describes the effect of the transformed ratio of algae to coral on short-term change, and the second column $\mathbf{a}_2$describes the effect of the transformed ratio of 'other' to the geometric mean of coral and algae. If $\mathbf{A} = \mathbf{0}$, then there are no such effects, and each component follows a stochastic exponential trajectory determined by $\mathbf{c}$.

The third term in Equation 1, $\boldsymbol{\beta}_1 \mathrm{z}_1(\mathrm{t})$, is the effect of the centred and scaled SST anomaly. Because $z_1(t)$ is centred and scaled, it has mean 0 and standard deviation 1. In consequence, $\boldsymbol{\beta}_1 \mathrm{z}_1(\mathrm{t})$ has mean vector $\mathbf{0}$ and covariance $\boldsymbol{\beta}_1 \boldsymbol{\beta}_1^{\mathrm{T}}$ (where the superscript T denotes transpose).

The fourth term in Equation 1, $\boldsymbol{\epsilon}_t$, is a multivariate normal error with mean vector $\mathbf{0}$ and constant covariance matrix $\boldsymbol{\Sigma}$. This includes the effects of processes on which we do not have data, such as storms.

A.5 Model fitting

Equation 1 is a VAR(1) (vector autoregressive model of order 1). We fitted Equation 1 using multivariate least squares (Lütkepohl 1993, pp. 62-65), implemented in the lm() function in R version 2.15.1 (R Core Team 2012). Multivariate least squares is consistent, and tests and confidence intervals based on the standard t and F statistics are asymptotically valid, even for nonstationary (Lütkepohl 1993, p. 369) or non-Gaussian (Hamilton 1994, p. 298) processes. Tests and confidence intervals are of course

approximate for finite samples. We used the Anova() function in the car package (version 2.0-15) to carry out multivariate hypothesis tests based on Pillai's trace, which may be more robust than other common multivariate test statistics (Hand and Taylor 1987). In each of these tests, we controlled for the effects of the other explanatory variables. For anomaly and climatology, we tested the hypotheses $\boldsymbol{\beta}_i = \mathbf{0}$ (that the explanatory variable had no effect on the expected transformed perturbing vector). For local threat levels, we tested the hypothesis $\boldsymbol{\Gamma} = \mathbf{0}$ (that there were no differences in expected transformed perturbing vectors between local threat levels). For reef composition, we tested the hypothesis $\mathbf{A} = \mathbf{0}$ (that the transformed composition had no effect on the expected transformed perturbing vector). For the local threat variables, we carried out post-hoc multivariate tests of the separate hypotheses $\boldsymbol{\gamma}_j = \mathbf{0}$ for $j = 1,2,3$ (that the effect of local threat levels Medium, High, and Very High did not differ from the effect of local threat level Low). Similarly, for reef composition, we carried out post-hoc multivariate tests of the separate hypotheses $\mathbf{a}_1 = \mathbf{0}$ and $\mathbf{a}_2 = \mathbf{0}$ (that the transformed ratio of algae to coral, and the transformed ratio of other to the geometric mean of coral and algae, had no effect on the expected transformed perturbing vector). Post-hoc tests were done using the linearHypothesis() function in the car package. We used Pillai's trace as the test statistic, and interpreted the resulting *P* values under a Bonferroni correction based on the number of post-hoc tests.

We concentrated on simple models without interactions, for which all of the parameters can be well estimated. We experimented with models including two- and three-way interactions, but for our data, these had some parameters that could not be identified, and others that had very large confidence ellipses and were very close to the boundary of the parameter space (which is often a symptom of overfitting). We also experimented with a multivariate mixed-effects model with a random intercept for each reef, which might account for some additional among-reef variability. We tried fitting this model in the R package sabreR version 1.1, but were unable to achieve satisfactory convergence.

Apart from the error covariance matrix $\mathbf{\Sigma}$, each of the parameters in the model (the intercept $\boldsymbol{\beta}_0$, the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and each column of $\mathbf{\Gamma}$ and $\mathbf{A}$) is a pair of isometric logratio coordinates, which can be back-transformed and plotted on a ternary plot. We added asymptotic 95% confidence ellipses as described in Tolosana-Delgado and van den Boogaart (2011). The error covariance matrix $\mathbf{\Sigma}$ can also be visualized on a ternary plot as an ellipse of unit Mahalanobis distance (the multivariate equivalent of one standard deviation) around the origin.

In addition to the stochasticity inherent in the model, the values of parameters are not known exactly. We used a nonparametric bootstrap with 10,000 pseudosamples to propagate this uncertainty.

The R code we used is available as supporting information.

A.6 Effects of observation error

We did not attempt to include observation error in our model. Although it is easy in principle to add observation error to models like Equation 1, it is difficult to estimate observation error from short time series, and we had only very short time series on individual reefs. It might be possible in future to estimate observation error across all the time series, assuming that observation error is drawn from the same distribution on each reef.

Instead, we established the likely consequences of observation error using simulations.

As described in Appendix A.1, each reef composition is estimated from the benthic organisms present at 200 points on each of 5 transects at each of 3 sites (Abdo et al. 2004). If we make the simplifying assumption that these points are independent, then we can treat the estimated composition as being drawn

from a multinomial with $3 \times 5 \times 200 = 3000$ observations. We generated 10000 simulated data sets from the fitted model, each with the addition of multinomial observation error, and refitted the model to each one. The estimated biases in the effects of SST anomaly, climatology, medium threat and very high threat were small (Fig. A3, green, pink, yellow and red crosses), as was the bias in the estimated covariance matrix (Fig A3, grey dotted line). There was a larger estimated bias in the effect of high threat (Fig. A3, orange cross), but this parameter also has a very large 95% confidence ellipse, so the bias is not likely to be accurately estimated. More importantly, the intercept and composition effects were all biased away from the origin (Fig. A3, grey, light blue and dark blue crosses). Because both the composition effects and the intercept affect the location of the stationary mean, biases in these parameters will also affect the estimated long-term effect of increased climatology. The strength of this long-term effect was underestimated in the presence of observation error, but the direction of the effect was approximately correct (Fig. A4, red dotted arrow).

A.7 Model checking

Our model assumes that the residuals are bivariate normal with constant covariance matrix, and do not systematically depend on explanatory variables or on time. We checked the assumption of bivariate normality using bivariate scatter plots of residuals, and identified outliers using the method described in Filzmoser et al. (2005) and implemented in the aq.plot() function in the R package mvoutlier, with default parameters. There were 30 potential outliers, mostly representing unusually large increases in other given the relative abundances of algae and coral (Fig. A5, red crosses). These outliers did not appear to be spatially segregated (Fig. A1, crosses). Disturbance events and transient dynamics may result in outliers,

but only if they cause year-to-year changes that are very different from the distribution of most of the data, after accounting for explanatory variables.

Deleting these 30 potential outliers had little effect on estimated parameters (Fig A6). We have two reasons to believe that our models are fairly robust. First, multivariate linear models with large sample sizes are robust to moderate departures from multivariate normality (Hand and Taylor 1987, p. 202). Second, an earlier analysis (Żychaluk et al. 2012) based on quite different assumptions (a semi-parametric model in which $\mathbf{y}(t+1)$ was assumed to have a Dirichlet distribution conditional on $\mathbf{y}(t)$, and no environmental variables were included) gave similar stationary distributions to those described below for current climatology. Biologically, outliers might arise either from heterogeneity among reefs, or because large changes in composition are more frequent than expected under a multivariate normal model, even on a single reef. Such heavy-tailed distributions have often been discussed in the context of population variability, although they may not be common in nature (Halley and Inchausti 2002).

We checked the assumptions of constant covariance and absence of systematic biases using scatter plots of predicted against observed values (Fig. A7). The only obvious problem was that when the observed year-to-year increase in algae was large, the predicted increase tended to be too small (Fig. A7B). This affects a relatively small number of observations, but has the potential to be biologically important.

We checked the assumption of no systematic dependence on explanatory variables or time using scatter plots of residuals against explanatory variables and time. There were no obvious patterns in plots of residuals against explanatory variables (Fig. A8).

There was a very slight tendency for our model to predict too little increase in coral and other, and too much increase in algae, in later years relative to earlier years (Fig. A9). Coral cover on the GBR is believed to have declined substantially over the last few decades (De'ath et al. 2012). The pattern we observed in residuals is in the opposite direction. This suggests that any decline in coral cover is being captured by the explanatory variables already in our model, and that changes in variables not included in our models may slightly mitigate this decline. In other words, if the trend in residuals was maintained, long-term projections from our model would be too pessimistic. We do not include this explicit temporal term in our analysis because it is very small relative to the effects in the model, and 10 years of data is not long enough to give a good idea of whether this is a genuine trend.

We investigated spatial autocorrelation in residuals by plotting multivariate spline correlograms as described in Bjornstad and Falck (2001). When pooled over years, there was only very weak spatial autocorrelation in residuals (Figure A10A). However, the pooled correlogram might hide temporally varying spatial autocorrelation resulting from events such as cyclones, bleaching and crown-of-thorns starfish. We therefore also examined correlograms for each year separately. In general, the 95% confidence envelopes for these correlograms (Figure A10B) were so wide that we have little information on possible year-by-year spatial autocorrelation.

A.8 Long-term behavior

Assume that lagged annual mean SST anomaly is a sequence of identically normally distributed random variables, independent of past SST anomalies and of the error term $\boldsymbol{\epsilon}(t)$. Then the two stochastic terms on the right-hand side of Equation 1 can be summed into a single multivariate normal random variable $\boldsymbol{\eta}(t) = \boldsymbol{\beta}_1 z_1(t) + \boldsymbol{\epsilon}(t)$, with mean vector $\mathbf{0}$ and covariance $\boldsymbol{\Sigma} + \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^{\mathrm{T}}$. Let $\mathbf{B} = \mathbf{A} + \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Then Equation 1 can be written in the form

$$\mathbf{x}(t+1) = \mathbf{c} + \mathbf{B}\mathbf{x}(t) + \boldsymbol{\eta}(t), \tag{A.4}$$

whose long-term behavior is well-known. Given a fixed initial vector $\mathbf{x}(0)$, it can be shown that at any subsequent time t, $\mathbf{x}(t)$ has a multivariate normal distribution with mean

$$E[\mathbf{x}(t)] = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^t)\mathbf{c} + \mathbf{B}^t\mathbf{x}(0) \tag{A.5}$$

and covariance

$$V[\mathbf{x}(t)] = \mathrm{vec}^{-1}\left[(\mathbf{I} - \mathbf{B} \otimes \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B}^t \otimes \mathbf{B}^t)\mathrm{vec}\left(\boldsymbol{\Sigma} + \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^{\mathrm{T}}\right)\right], \tag{A.6}$$

where $\otimes$ is the Kronecker product and vec() is the operator that stacks the columns of a matrix into a single column (Harville 2008, chapters 16 and 21). If $\mathbf{B}$ has spectral radius (absolute value of largest eigenvalue) less than 1, then the terms in Equations A.5 and A.6 involving $\mathbf{B}^t$ go to zero as $t \to \infty$, and $\mathbf{x}(t)$ approaches a multivariate normal stationary distribution with mean vector

$$\boldsymbol{\mu}^* = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{c} \tag{A.7}$$

and covariance

$$\boldsymbol{\Sigma}^* = \mathrm{vec}^{-1}\left[(\mathbf{I} - \mathbf{B} \otimes \mathbf{B})^{-1}\mathrm{vec}\left(\boldsymbol{\Sigma} + \boldsymbol{\beta}_1 \boldsymbol{\beta}_1^{\mathrm{T}}\right)\right] \tag{A.8}$$

(Hamilton 1994, chapter 10).

Equations A.4 to A.8 give several biological insights. First, from Equations A.5 and A.6, the existence of a stationary distribution whose properties are independent of initial conditions depends only on $\mathbf{B}$, the

term in the model describing the effects of reef composition on dynamics. If the spectral radius of **B** is less than 1, then the effects of any disturbance to $\mathbf{x}(t)$ in Equation A.4 will die away over time, rather than being amplified, and a stationary distribution exists.

Second, if the spectral radius is less than 1, then the smaller its value, the more quickly the terms in Equations A.5 and A.6 involving $\mathbf{B}^t$ go to zero, and thus the more quickly the stationary distribution is approached. In this model, the amount of variability has no effect on the existence or otherwise of a stationary distribution. If the eigenvalues of **B** are complex, then the approach to the stationary distribution will involve oscillations.

Third, if the stationary distribution exists, then its expected value in ilr coordinates (Equation A.7) depends on both **B** (which describes the effects of reef composition on dynamics) and **c** (which describes the deterministic aspect of reef dynamics in the absence of effects of reef composition).

Fourth, if the stationary distribution exists, its covariance (Equation A.8) depends on **B** (which describes the effects of reef composition on dynamics), **Σ** (the covariance matrix for stochastic effects other than fluctuations in lagged annual mean SST anomaly), and $\boldsymbol{\beta}_1$ (the effect of fluctuations in lagged annual mean SST anomaly). In this model, increasing variability, either in SST anomaly or in other stochastic effects, can increase the spread of the stationary distribution, but cannot alter the location of the stationary mean in ilr coordinates.

The stationary distribution in the original sample space is a logistic normal distribution, obtained by back-transformation. If the multivariate normal has density $f(\mathbf{x})$ at point $\mathbf{x}$, then the density at the corresponding composition $\mathbf{y} = \text{ilr}^{-1}(\mathbf{x})$ is $f(\mathbf{y}) = \frac{1}{\sqrt{3}y_1y_2y_3} f(\mathbf{x})$, where ilr^{-1} denotes the inverse of the isometric logratio transformation (Monti et al. 2011). Back-transformation has two important consequences. First, although the stationary distribution is unimodal in ilr coordinates, the back-transformed stationary distribution can have one, two, or three modes (e.g. Figure 3.3b in Mateu-Figueras et al. 2011), and is therefore capable of describing distributions resulting from systems in which there is more than one likely set of reef compositions, separated by compositions that are less likely. Such patterns are one stochastic analogue of the concept of alternative stable states in a deterministic system. Second, the back-transformed mean vector $\boldsymbol{\mu}^*$ is known as the metric centre of the distribution, and is the preferred measure of location for a compositional distribution (Aitchison 1989). It does not in general coincide with the arithmetic mean of the distribution, whose location will be influenced by the stationary covariance $\boldsymbol{\Sigma}^*$ as well as the stationary mean vector $\boldsymbol{\mu}^*$.

If $\mathbf{B}$ has spectral radius 1 or greater, then the terms involving $\mathbf{B}^t$ in Equations A.5 and A.6 do not go to zero as $t \to \infty$. The model does not have a stationary distribution in such cases, giving another analogue to alternative stable states in a deterministic system. The long-term behavior of the expected value depends on the initial condition $\mathbf{x}(0)$. The covariance also continues to change over time, although it does not depend on the initial condition. We do not study such cases in detail, because there was very little support for them in the data (as evaluated by the bootstrap described in Appendix A.10). Instead, we concentrate on the properties of the stationary distribution, and report the proportion of bootstrap replicates in which $\mathbf{B}$ had spectral radius less than 1.

Probability of undesirable reef compositions can be found by numerical integration, given the parameters of the stationary distribution.

A.9 Sensitivity of the stationary distribution and climatology equivalent of local threat effects

Differentiating the stationary mean (Equation A.7) with respect to mean climatology u_2 gives

$$\frac{\mathrm{d}\boldsymbol{\mu}^*}{\mathrm{d}u_2} = (\mathbf{I} - \mathbf{B})^{-1} \frac{1}{s_2} \boldsymbol{\beta}_2. \qquad \text{(A.9)}$$

The long-term consequence of the interplay between short-term climatology effects and population dynamics can be understood by comparing Equations A.9 and A.3. In the short term (Equation A.3), the derivative of the expected short-term change with respect to climatology is in the direction given by the climatology coefficient vector $\boldsymbol{\beta}_2$ (Fig. 5, solid arrow). In the long term (Equation A.9), the direction of the derivative of the stationary mean reef composition with respect to climatology is modified by the matrix $(\mathbf{I} - \mathbf{B})^{-1}$, which accounts for the way in which a short-term increase in algae is modified by reef composition in all successive years (Fig. 5, dashed arrow). Note that the only case in which the long-term effect of a change in climatology will be in the same direction as the short-term effect is if $\mathbf{c}$ is a right eigenvector of $(\mathbf{I} - \mathbf{B})^{-1}$ (or equivalently, of $\mathbf{A}$). Since almost all vectors are not eigenvectors of a given matrix, in almost all cases, short-term and long-term effects will be in different directions.

The derivative of the multivariate normal density $f(\mathbf{x})$ at composition $\mathbf{x}$ with respect to u_2 is

$$\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}u_2} = f(\mathbf{x})(\mathbf{x} - \boldsymbol{\mu}^*)^{\mathrm{T}} (\boldsymbol{\Sigma}^*)^{-1} \frac{\mathrm{d}\boldsymbol{\mu}^*}{\mathrm{d}u_2}. \qquad \text{(A.10)}$$

Equation A.10 evaluated at $u_2 = 0$ is the sensitivity of the stationary density to mean climatology. This sensitivity can be back-transformed to the original sample space in the same way as the density.

We used a similar approach to calculate the equivalent long-term effect of a local threat level in terms of increase in climatology. From Equation A.9, the long-term effect of a 1˚C increase in climatology is $\mathbf{d} = (\mathbf{I} - \mathbf{B})^{-1}\frac{\mathbf{1}}{s_2}\boldsymbol{\beta}_2$. Similarly, the long-term effect of the difference between low local threat level and some other level *j* is given by $\mathbf{e} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\gamma}_\mathbf{j}$.The projection of **e** onto the direction of **d** is the component of the local threat effect acting in the same direction as the climatology effect, and the norm of this projection ($\mathbf{e} \cdot \mathbf{d}/(\|\mathbf{d}\|^2)$) measures the long-term equivalent of the local threat effect in terms of degrees climatology increase.

A.10 Convergence to a stationary distribution

More than 98% of bootstrap replicates had spectral radius of **B** less than 1, and therefore converged to a stationary distribution. The bootstrap mean absolute value of the largest eigenvalue of **B** was 0.95 (bootstrap standard deviation 0.03), fairly close to the boundary beyond which such a distribution would not exist. No bootstrap replicates had complex eigenvalues, so there was no evidence for oscillations on approach to the stationary distribution.

References

Abdo, D., S. Burgess, G. Coleman, and K. Osborne. 2004. Surveys of benthic reef communities using underwater video. Long-term monitoring of the Great Barrier Reef Standard Operational Procedure Number 2. Townsville, Australia.

Aitchison, J. 1986. The statistical analysis of compositional data. The Blackburn Press, Caldwell, NJ.

Aitchison, J. 1989. Measures of location of compositional data sets. Mathematical Geology **21**:787-790.

Bjornstad, O. N., and W. Falck. 2001. Nonparametric spatial covariance functions: Estimation and testing. Environmental and Ecological Statistics **8**:53-70.

Burke, L., K. Revtar, M. Spalding, and A. Perry. 2011. Reefs at Risk Revisited. World Resources Institute, Washington, DC.

De'ath, G., K. E. Fabricius, H. Sweatman, and M. Puotinen. 2012. The 27-year decline of coral cover on the Great Barrier Reef and its causes. Proceedings of the National Academy of Sciences of the United States of America **109**:17995-17999.

Filzmoser, P., R. G. Garrett, and C. Reimann. 2005. Multivariate outlier detection in exploration geochemistry. Computers & Geosciences **31**:579-587.

Halley, J., and P. Inchausti. 2002. Lognormality in ecological time series. Oikos **99**:518-530.

Hamilton, J. D. 1994. Time series analysis. Princeton University Press, Princeton, New Jersey.

Hand, D. J., and C. C. Taylor. 1987. Multivariate analysis of variance and repeated measures: a practical approach for behavioural scientists. Chapman and Hall, London.

Harville, D. A. 2008. Matrix algebra from a statistician's perspective. Springer, New York.

Lütkepohl, H. 1993. Introduction to multiple time series analysis. Springer-Verlag, Berlin.

Mateu-Figueras, G., V. Pawlowsky-Glahn, and J. J. Egozcue. 2011. The principle of working on coordinates. Pages 31-42 *in* V. Pawlowsky-Glahn and A. Buccianti, editors. Compositional data analysis: theory and applications. John Wiley & Sons, Chichester.

Monti, G. S., G. Mateu-Figueras, and V. Pawlowsky-Glahn. 2011. Notes on the scaled Dirichlet distribution. Pages 129-138 *in* V. Pawlowsky-Glahn and A. Buccianti, editors. Compositional data analysis: theory and applications. John Wiley & Sons, Chichester.

R Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Selig, E. R., K. S. Casey, and J. F. Bruno. 2010. New insights into global patterns of ocean temperature anomalies: implications for coral reef health and management. Global Ecology and Biogeography **19**:397-411.

Tolosana-Delgado, R., and K. G. van den Boogaart. 2011. Linear models with compositions in R. Pages 356-371 *in* V. Pawlowsky-Glahn and A. Buccianti, editors. Compositional data analysis: theory and applications. John Wiley & Sons, Chichester.

van den Boogaart, K. G., and R. Tolosana-Delgado. 2008. "compositions": A unified R package to analyze compositional data. Computers & Geosciences **34**:320-338.

Żychaluk, K., J. F. Bruno, D. Clancy, T. R. McClanahan, and M. Spencer. 2012. Data-driven models for regional coral-reef dynamics. Ecology Letters **15**:151-158.

Appendix tables

Table A1. AIC values for alternative models.

Model	edf	AIC
Lagged SST anomaly, lagged climatology, local threat, composition	372	81.4
Lagged SST anomaly, lagged climatology, distance from coast, composition	370	97.7
Lagged SST anomaly, local threat, composition	371	97.8
Lagged SST, local threat, composition	371	99.9
Lagged SST, distance from coast, composition	369	110.2
Lagged SST anomaly, distance from coast, composition	369	115.9
SST anomaly, local threat, composition	371	146.1
Lagged climatology, local threat, composition	371	148.6
Climatology, local threat, composition	371	150.4
Lagged climatology, distance from coast, composition	369	166.5
SST anomaly, distance from coast, composition	369	166.9
Climatology, distance from coast, composition	369	168.9
SST, local threat, composition	371	169.0
SST, distance from coast, composition	369	190.1

Notes: edf (effective degrees of freedom) and AIC calculated using `extractAIC()` in R.

Table A2. Effects of reef composition, lagged sea surface temperature anomaly and climatology and local threat on short-term change in GBR composition.

Term	df	Pillai trace	approx *F*	num df	den df	*P*
Composition	2	0.19	18.25	4	712	3×10^{-14}
SST anomaly	1	0.11	21.00	2	355	2×10^{-9}
SST climatology	1	0.03	4.98	2	355	0.007
Local threat	3	0.06	3.92	6	712	0.0007

Notes: each term is tested by comparing a model containing all terms to a model without the term of interest, which has df fewer parameters. The null distribution of the resulting Pillai trace statistic is approximately an *F* distribution with (num df, den df) degrees of freedom, which was used to obtain the *P* value. The composition term is the **A** matrix in Equation 1.

Table A3. Post-hoc tests for terms in the model for short-term change.

Term	df	Pillai trace	approx *F*	num df	den df	*P*
$\mathbf{a}_1$	1	0.05	8.94	2	355	0.0001
$\mathbf{a}_2$	1	0.08	15.27	2	355	4×10^{-7}
medium local threat	1	0.05	9.30	2	355	0.0001
high local threat	1	0.009	1.69	2	355	0.19
very high local threat	1	0.003	0.48	2	355	0.62

Notes: $\mathbf{a}_1$ (the first column of the **A** matrix in Equation 1) describes the effect of the transformed ratio of algae to coral on the transformed perturbing vector, and $\mathbf{a}_2$ (the second column of **A**) describes the effect of the transformed ratio of other to the geometric mean of coral and algae on the transformed perturbing vector. Each of the local threat terms tests the hypothesis that the effect of that local threat level does not differ from that of low local threat. *P* values should be interpreted under a Bonferroni correction for five tests, so that, for example, a *P* value of 0.01 would be required for significance at the 0.05 level.

Appendix figures

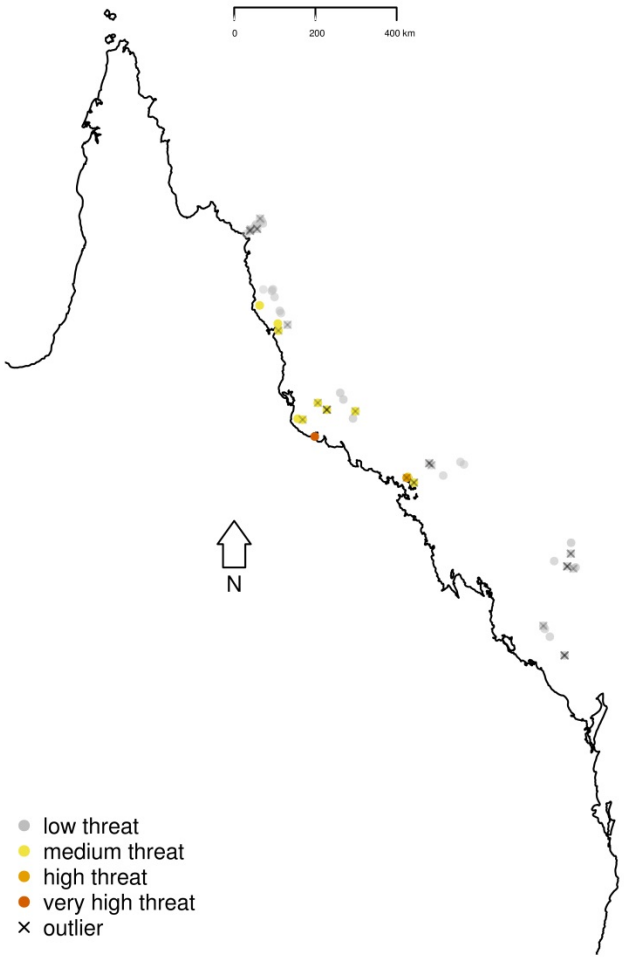

Fig. A1. Locations of reefs (circles) colour-coded by local threat level. Locations of potential outliers are shown as crosses.

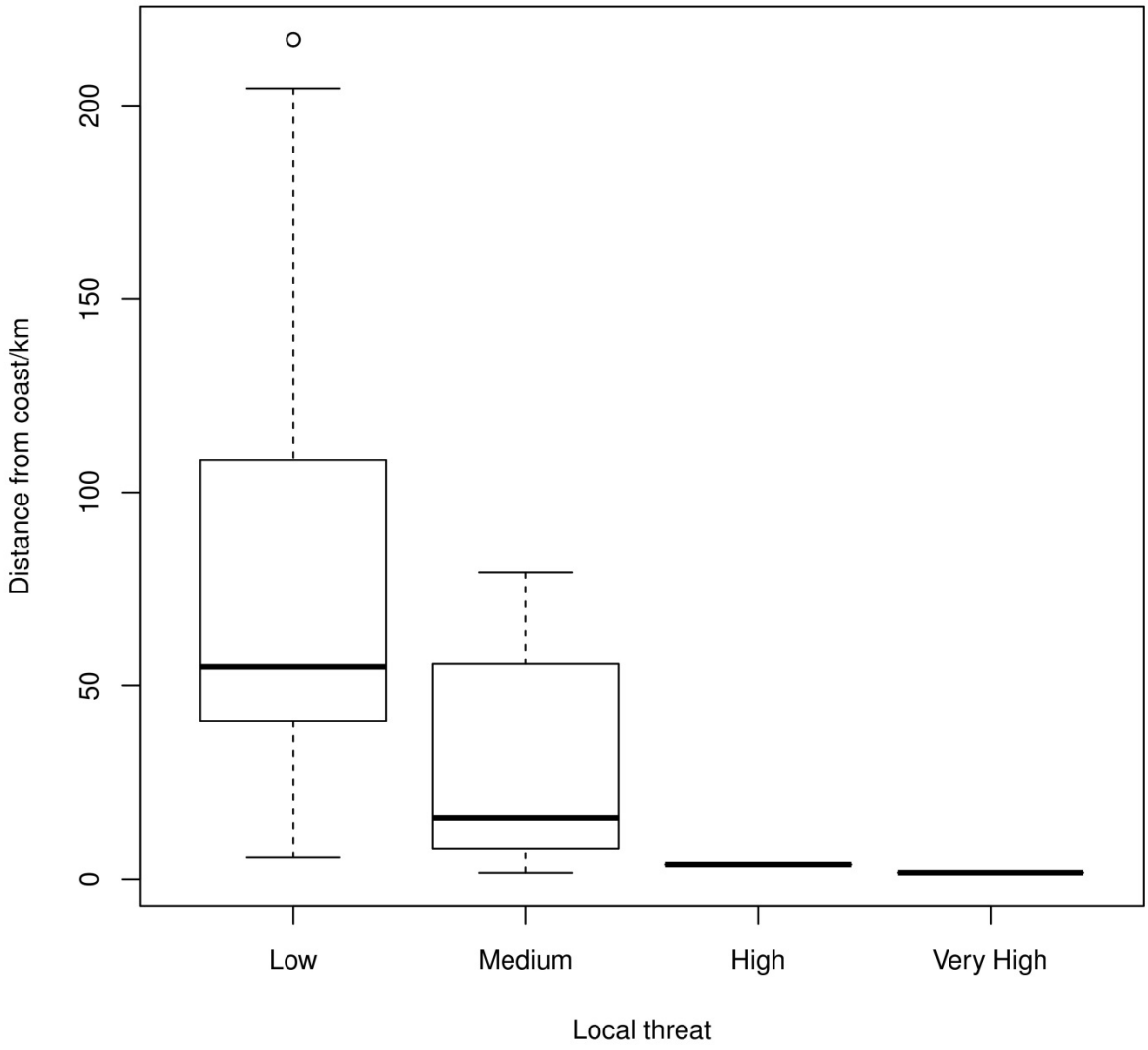

Fig. A2. Relationship between shortest distance from reef to coast in kilometres and local threat index.

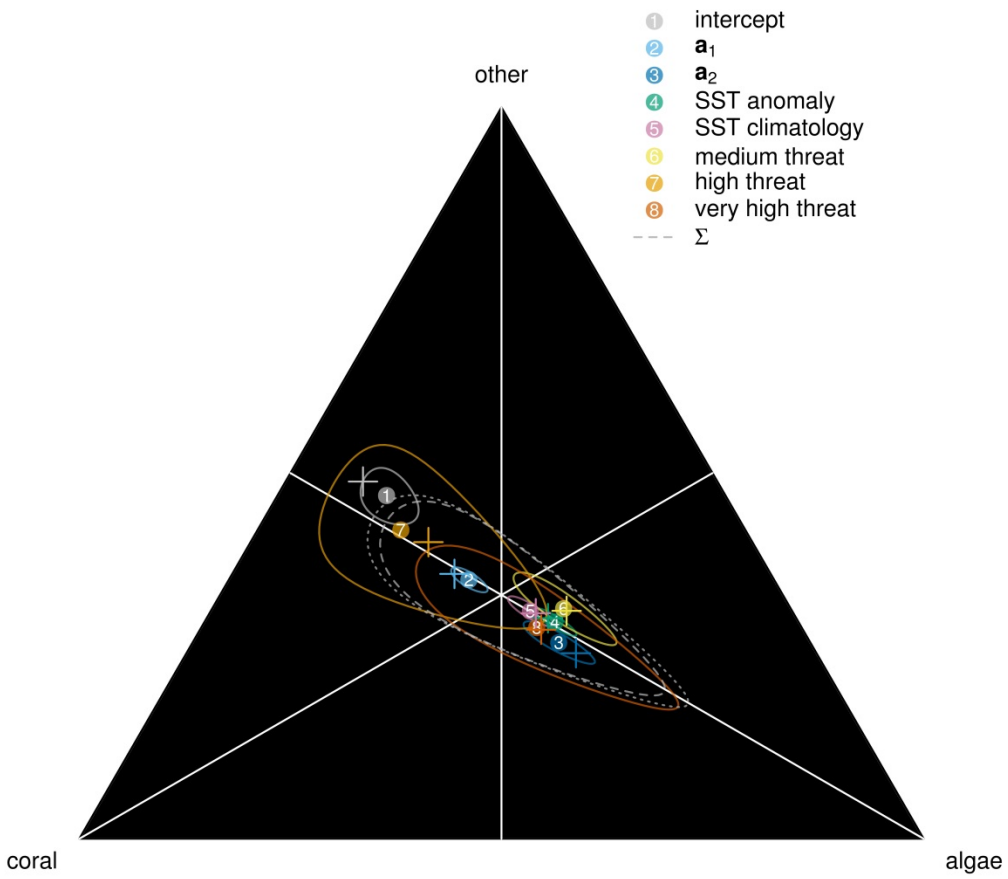

Fig. A3. Effects of simulated observation error on parameter estimates. Circles are true parameter values (the estimates from Fig. 1C). Crosses with corresponding colours are mean estimates from 10000 replicate simulated data sets generated under the fitted model, with the addition of multinomial observation error (3000 points). Grey dashed line: shape of the true covariance matrix $\mathbf{\Sigma}$, represented by an ellipse at unit Mahalanobis distance around the no-effect point. Grey dotted line: mean estimated shape of covariance matrix.

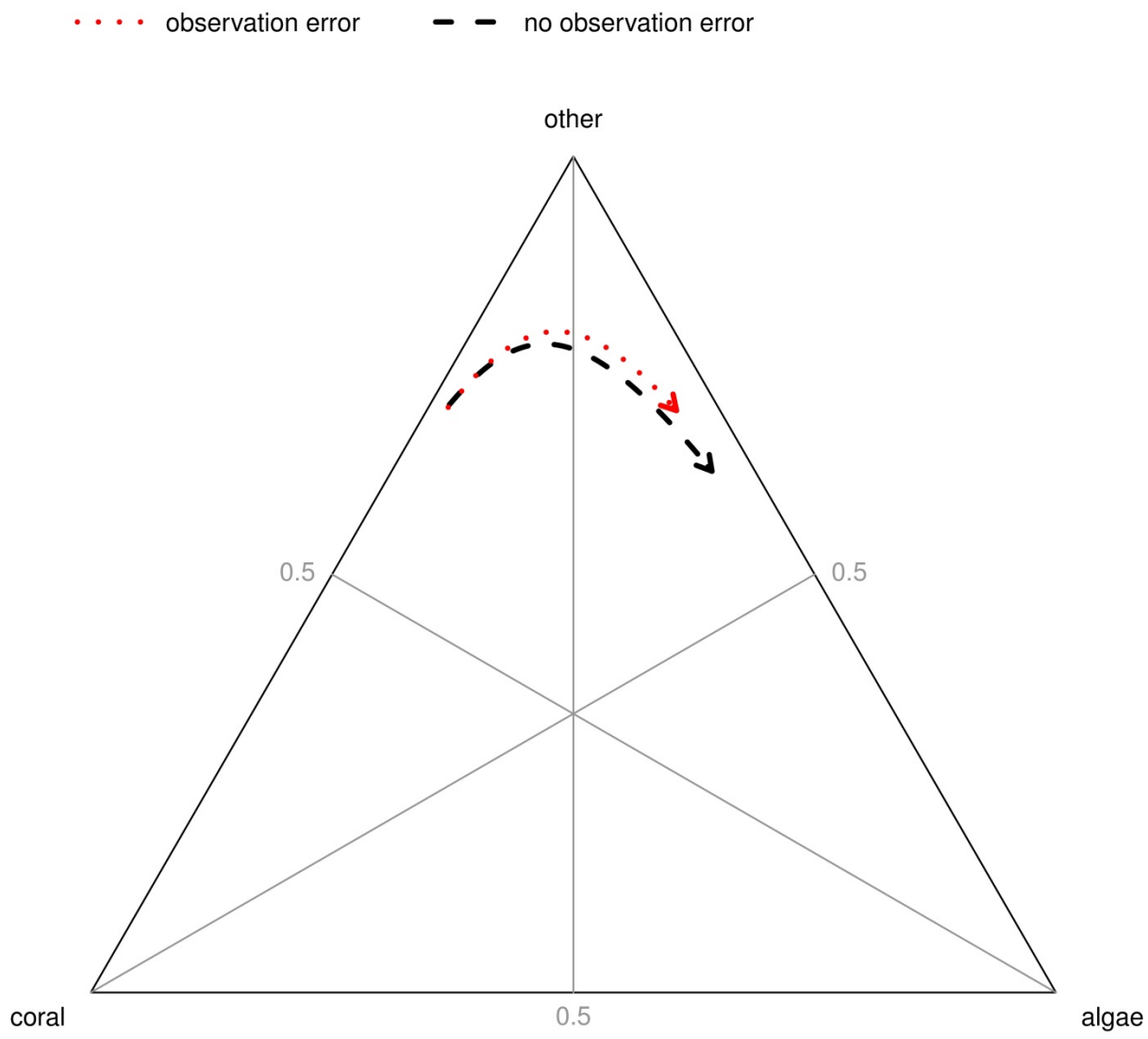

Fig. A4. Estimated long-term direction of increased climatology (scaled to 3.5°C increase) under the fitted model without observation error (black dashed arrow, as in Fig. 5), and mean estimated long-term direction from 10000 simulated data sets with observation error as in Fig. A3 (red dotted arrow).

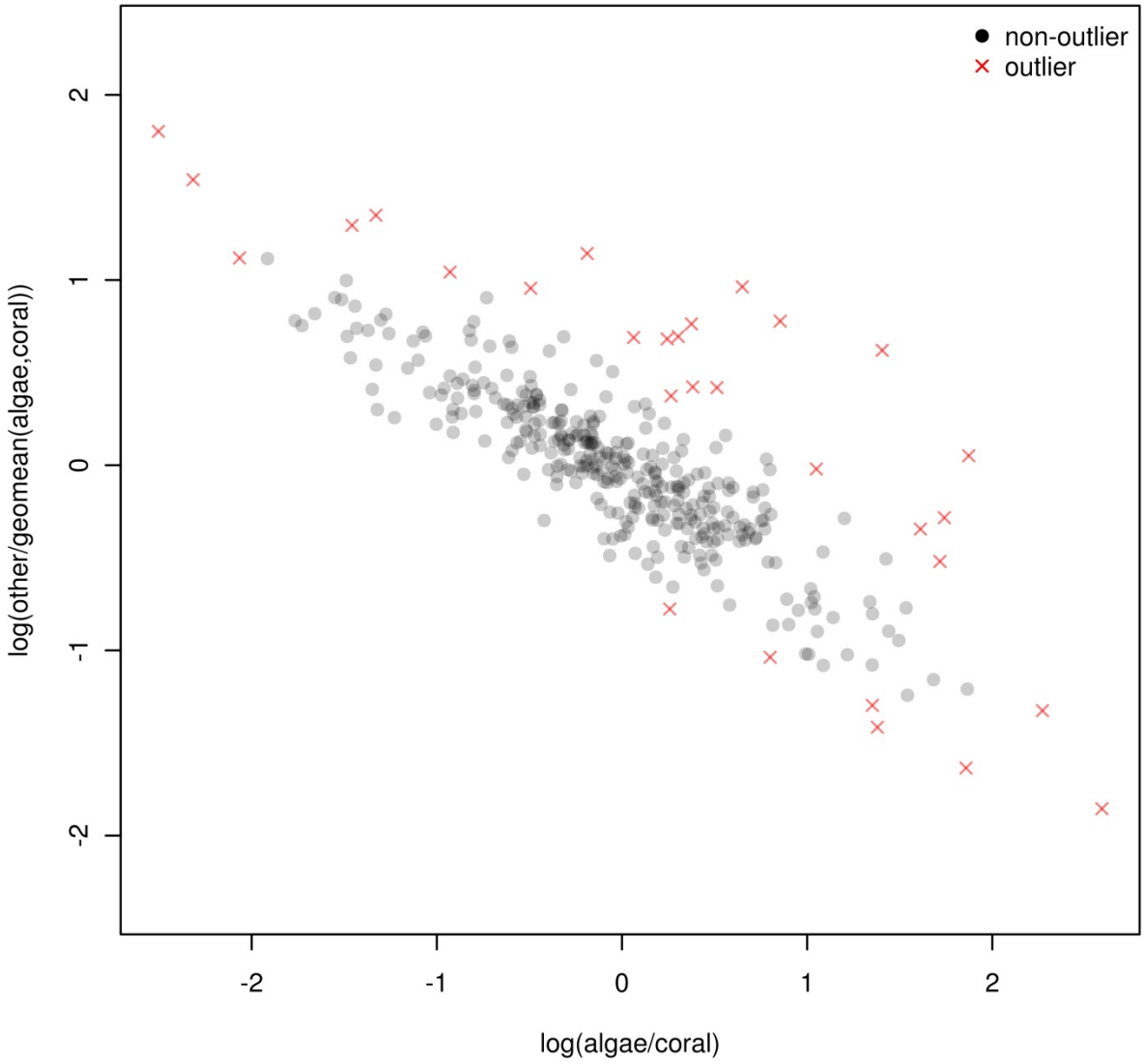

Fig. A5. Scatterplot of residuals, with red crosses indicating potential outliers.

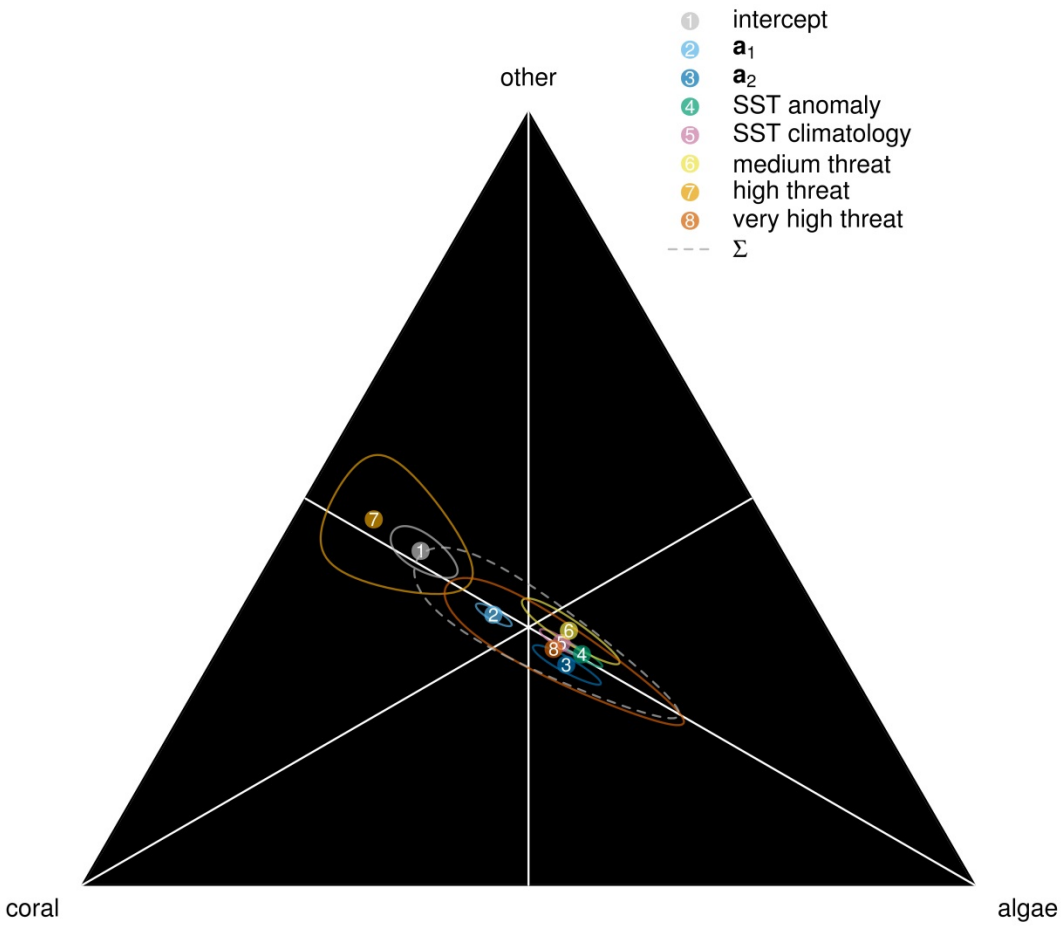

Fig. A6. Estimated parameters after the deletion of 30 potential outliers. Symbols as in Fig. 1C.

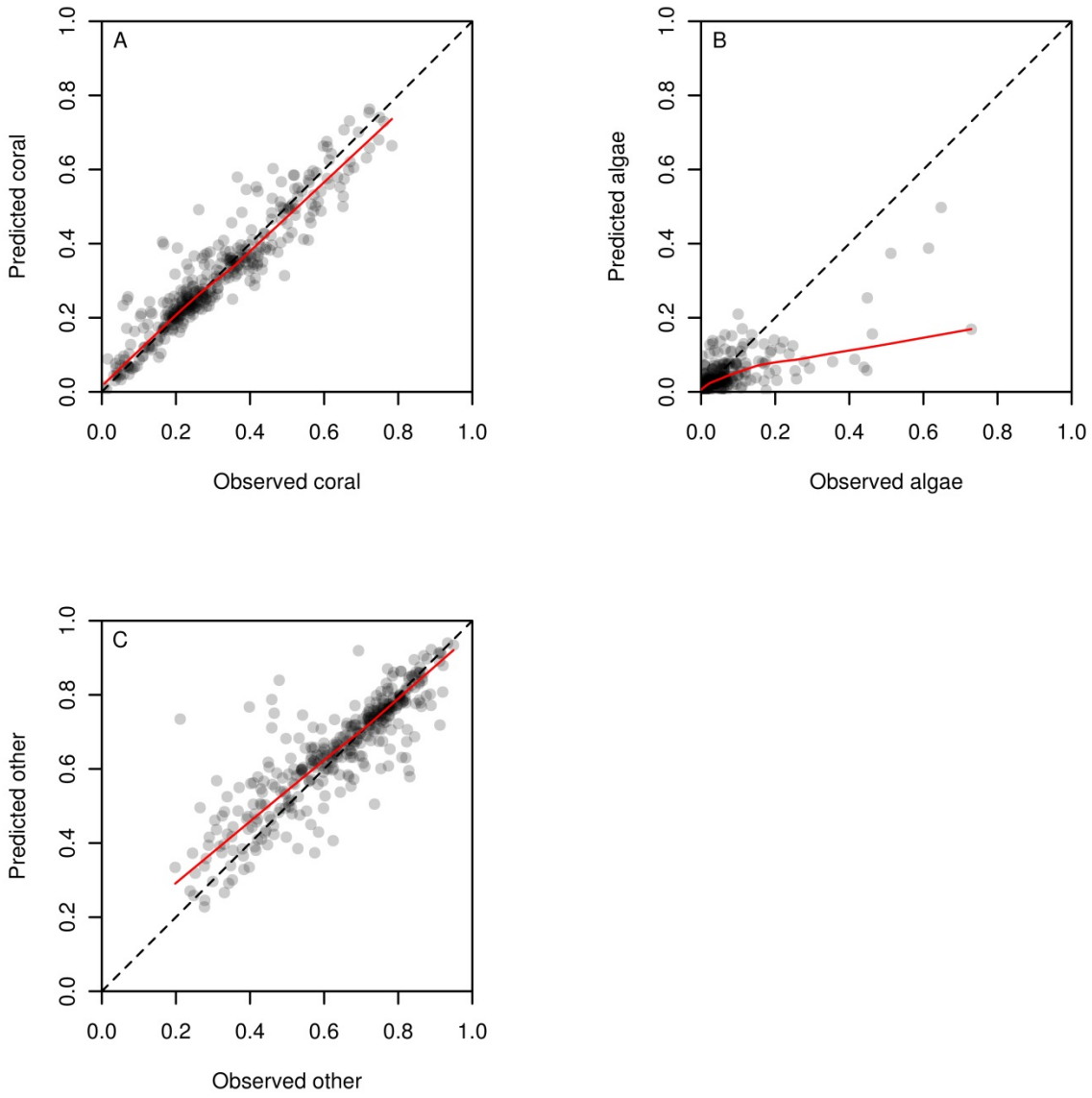

Fig. A7. Scatterplots of predicted against observed year-to-year changes in reef composition (circles) with lowess smoothers (red lines) and 1:1 lines (black dashed lines).

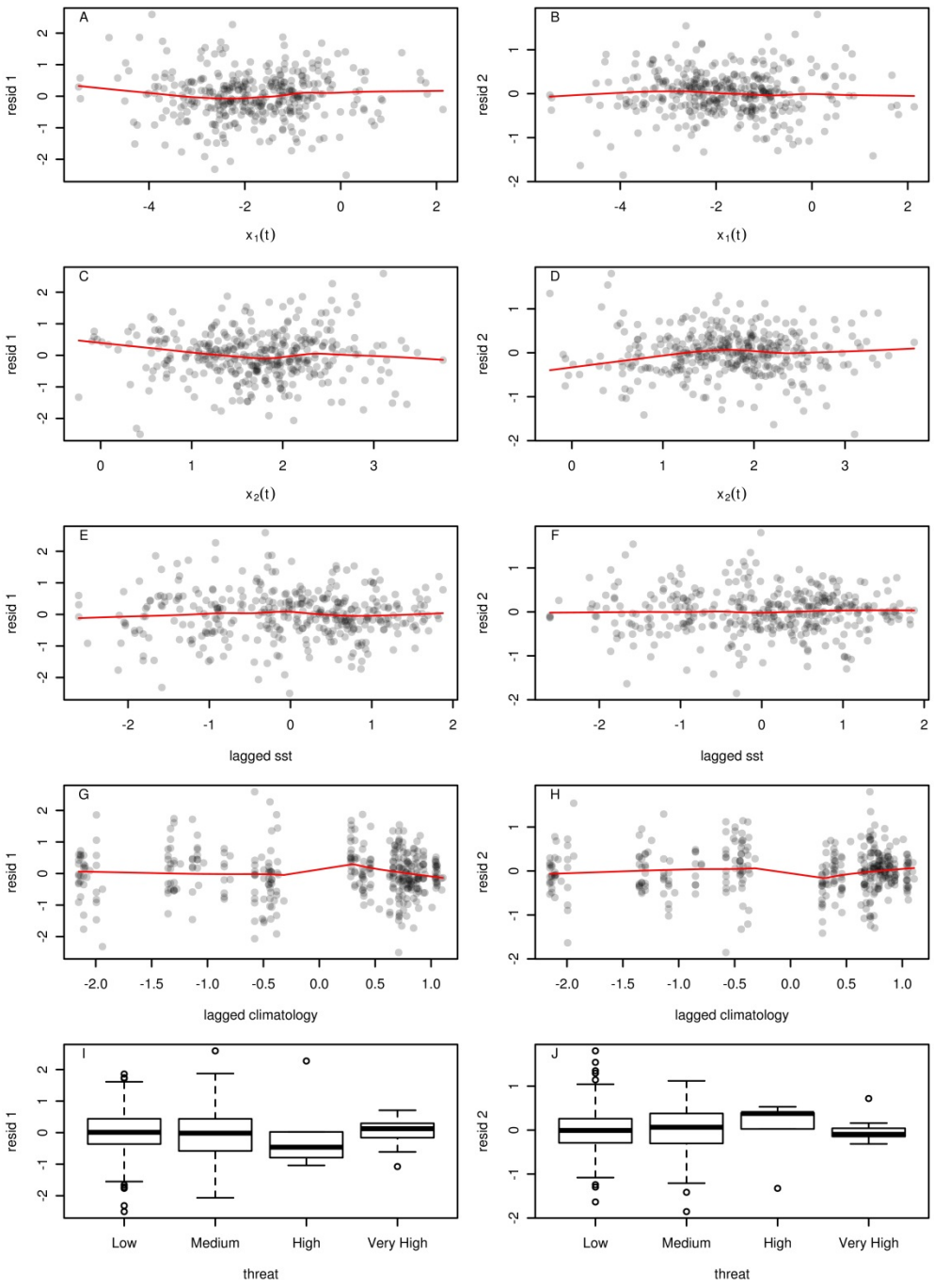

Fig. A8. Plots of residuals against explanatory variables. The y-axis variable resid 1 in A, C, E, G, I is residual year-to-year change in log(algae/coral). The y-axis variable resid 2 in B, D, F, H, J is residual year-to-year change in log(other/geomean(algae, coral)). In A-D, $x_1(t)$ and $x_2(t)$ are the two components of ilr-transformed composition at time t. Red lines are lowess smoothers.

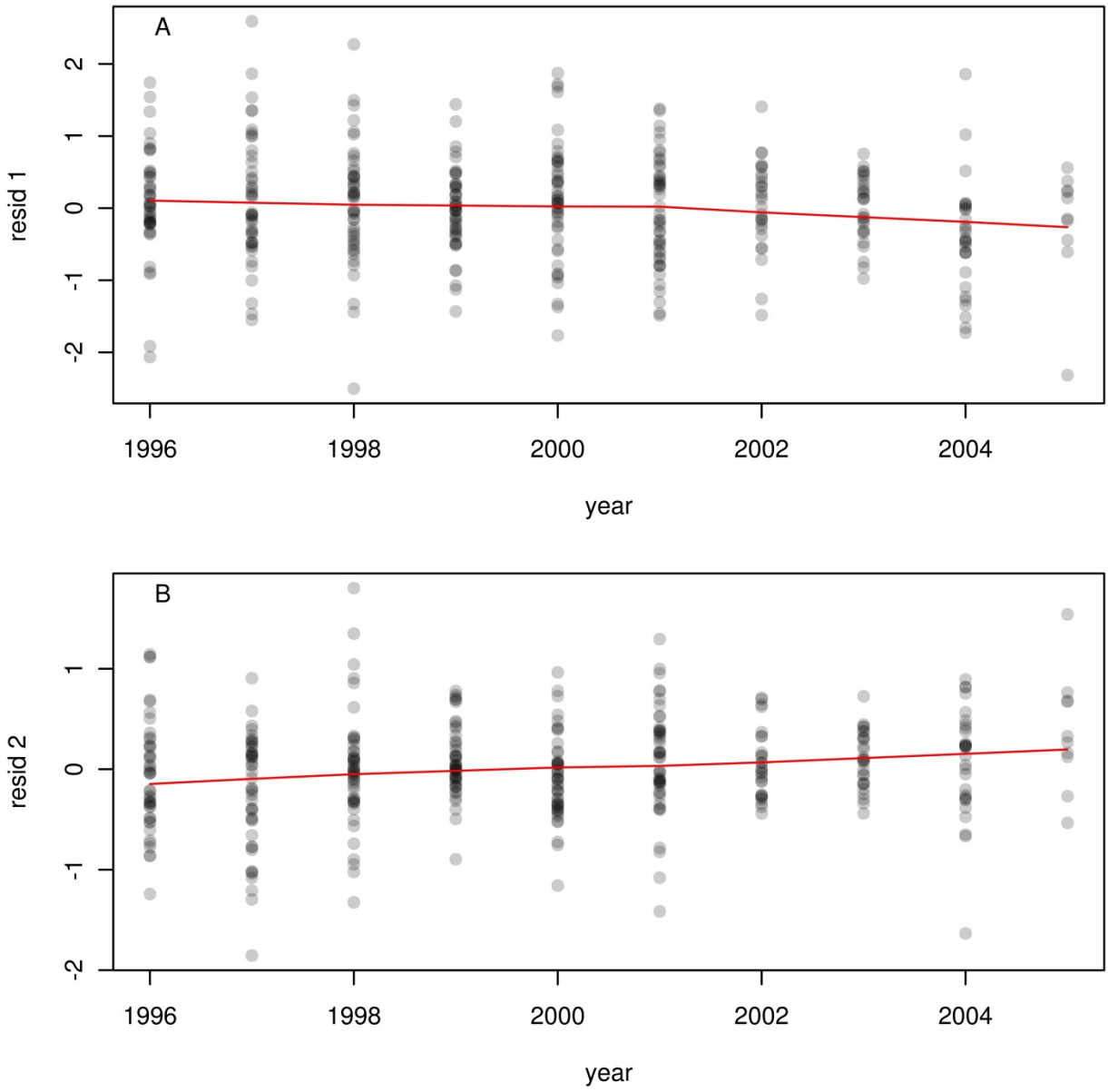

Fig. A9. Plots of residuals against time. The y-axis variable resid 1 in A is residual year-to-year change in log(algae/coral). The y-axis variable resid 2 in B is residual year-to-year change in log(other/geomean(algae, coral)). Red lines are lowess smoothers.

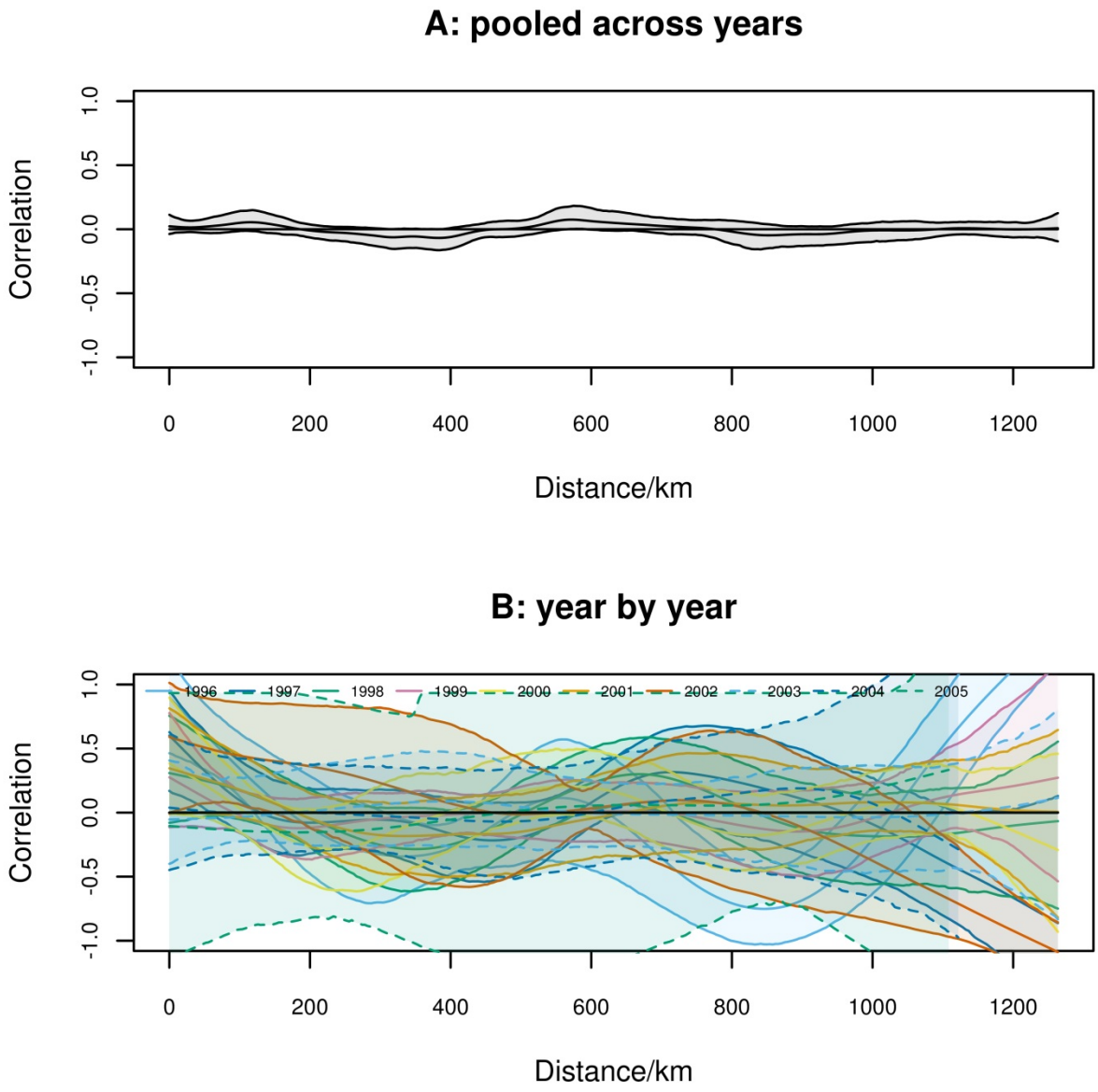

Figure A10. Spatial autocorrelation of residuals. A: pooled across all years. B: year by year. Multivariate spline correlograms (Bjornstad and Falck 2001) calculated using the `spline.correlog()` function in the R package `ncf` version 1.1-5 with default spline degrees of freedom. Shaded regions are 95% pointwise envelopes, calculated from 1000 bootstrap resamples.

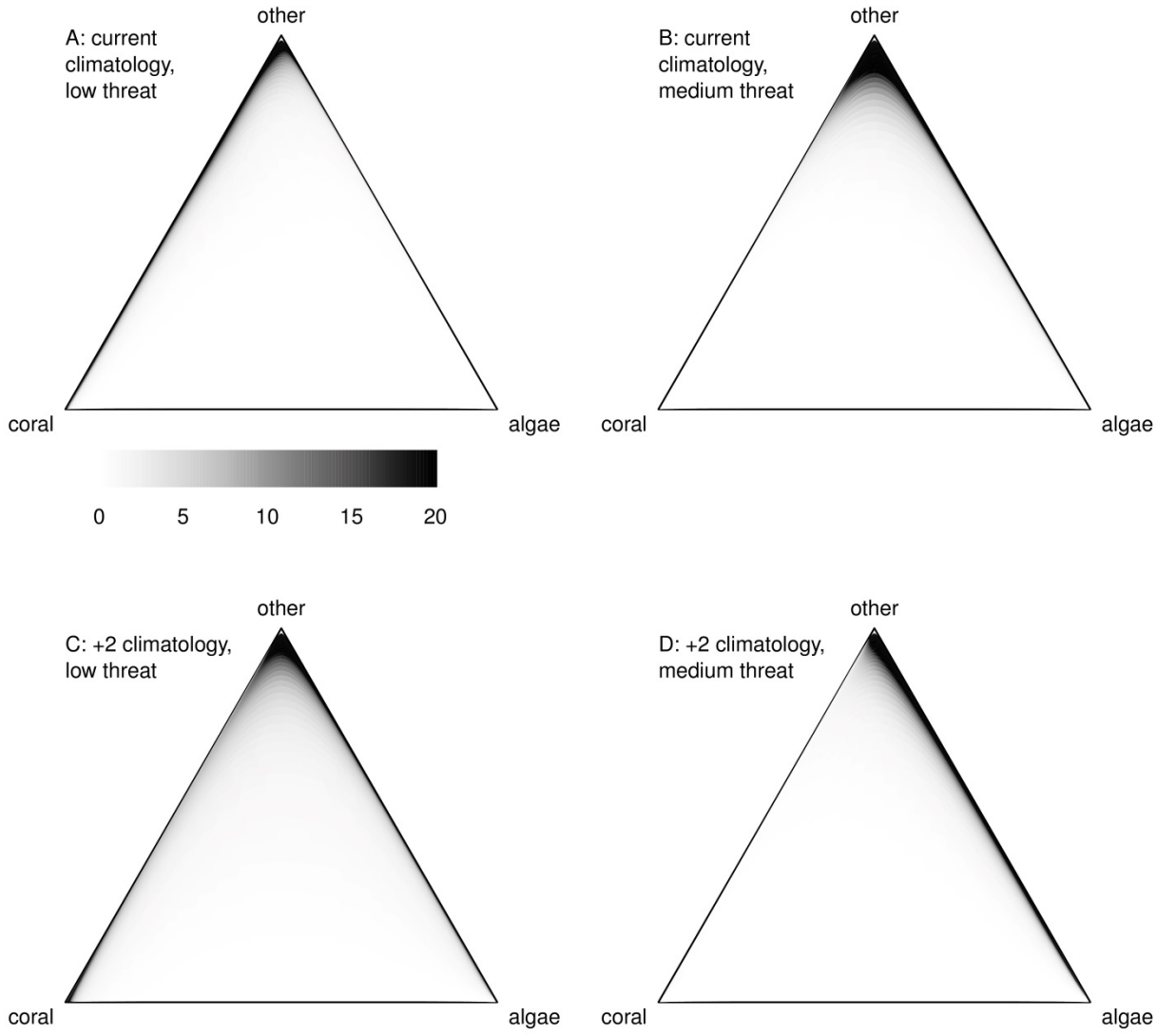

Figure A11. Bootstrap standard deviations of stationary density at current climatology (A: low local threat, B: medium local threat), and with a 2°C increase in mean climatology (C: low local threat, D: medium local threat). Darker colours are higher standard deviations.

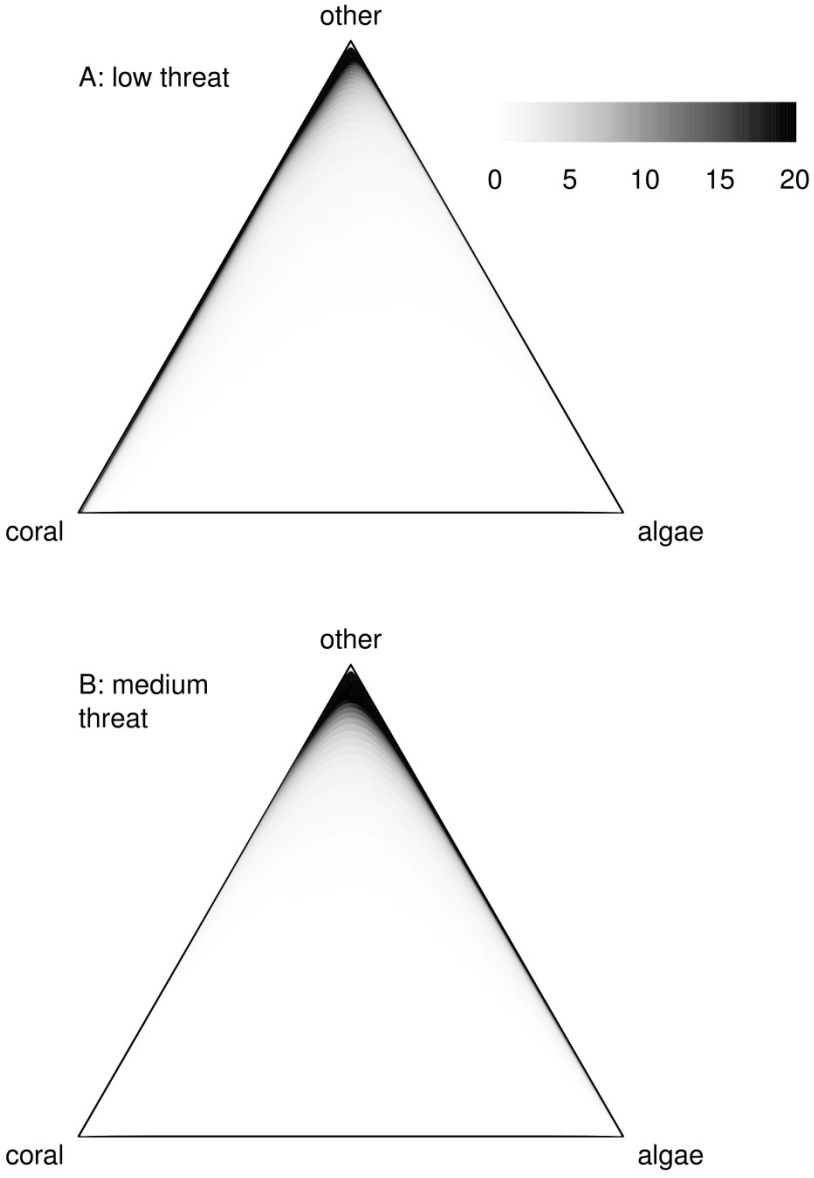

Figure A12. Bootstrap standard deviations of sensitivity of stationary density for the GBR to increases in climatology, evaluated at current mean climatology, and either low (A) or medium (B) local threat.