Data-driven models for regional coral-reef dynamics: supporting information

Kamila Żychaluk, John Bruno, Damian Clancy, Tim McClanahan, and Matthew Spencer

November 11, 2011

S1 Supplementary methods

S1.1 Data sources

Caribbean and Great Barrier Reef data are described in detail in Appendix A of Bruno et al. (2009). Data sources included published primary (peer-reviewed) literature, gray literature and unpublished survey data from governmental and non-governmental organization scientists and, for one source (Reef Check), volunteers trained and supervised by professional scientists. Quantitative survey data were collected in situ using SCUBA on fore reef environments between 1 and 15 m depth (mean depth $\sim$ 7 m). Surveys measured the percentage of the substratum covered by living coral and fleshy or calcareous macroalgae, primarily using some variant of the line-transect technique, in which a transect (typically a 10 to 30 m measuring tape or chain) was placed on the reef. The coverage of coral and macroalgae was then estimated either in situ by recording the number of points along each transect that overlaid corals, macroalgae, etc. or by taking images of the reef substrate at these points, which were then analyzed in the laboratory. All surveys differentiated macroalgae from other algal groups. Following Steneck (1988) and others, we defined macroalgae (i.e., seaweed) as "larger (canopy heights usually $>$ 10 mm), more rigid and anatomically complex algal forms". This functional group includes erect calcifying species (e.g. *Halimeda spp.*) but does not include microalgae and filamentous algae (turfs) or crustose algae (Steneck, 1988). Replicate cover measurements taken at different stations or depths at a given location were pooled into a single mean estimate for each reef in each year. Observations were made throughout the year, but we ignore seasonal variation for simplicity. We retained only those observations forming sequences of at least two observations on the same reef in successive calendar years (Caribbean: 69 reefs, median 2 observations per reef, range 2-7, covering the years 1997-2006. Great Barrier Reef: 55 reefs, median 9 observations per reef, range 2-11, covering the years 1996-2006).

Portions of the Kenyan data are described in McClanahan (2008). Reefs were sampled using the line-intercept technique (9 to 12 $\times$10 m transect lines) following the contour of the reef between 0.5 and 2 m depth at low tide. All sites are lagoon/backreef sites in Kenya. In Kenya, two locations at each of six reefs were sampled. At additional two reefs a single location was sampled. Surveys were performed at roughly 12-month intervals; the earliest in 1991 and the latest in 2009, although some reefs were not sampled for the entire period. Calendar year 1998 is skipped because samples were taken in December up until calendar year 1997, then January from calendar year 1999. We retained consecutive sequences of at least two observations on the same location at roughly 12-month intervals (median 18 observations per location, two locations observed 5 times, others observed 16-18 times).

A Google Earth kml file showing the locations of study sites is available from `http://www.liv.ac.uk/~matts/coralsimplex.html`.

S1.2 Assumptions

Here, we give more explanation of the assumptions outlined in the main text.

1. Reefs within a region are independent of each other. We ignore spatial autocorrelation caused by processes that affect multiple reefs, e.g. large storms, and by dispersal (although we do not require the assumption that each reef is a closed system). This is unlikely to be strictly true (Ninio et al., 2000; Cheal et al., 2007), but we do not have enough data to model the

dependencies. Furthermore, although reefs are linked by dispersal, larval recruitment may not be limiting, at least in many parts of the Great Barrier Reef (Hughes et al., 1999; van Woesik et al., 1999; Wooldridge et al., 2005). Additionally, many studies have noted the seemingly independent temporal dynamics of neighbouring reefs (Edmunds and Bruno, 1996; Murdoch and Aronson, 1999) and the highly localized effects of even large scale disturbances including storms, disease outbreaks and bleaching events (Bythell et al., 2000; Berkelmans et al., 2004; Bruno and Selig, 2007). Overall, we think that dependence between reefs is likely to be weak, and it has been shown that methods similar to ours perform well under weak dependence (e.g. Masry and Fan, 1997; Hallin et al., 2004).

Furthermore, although it is possible that spatial autocorrelation reduces our effective sample size, this would increase rather than decrease our ability to detect alternative stable states if they existed. We select the proportion of reefs to include in the set of neighbours of a given state using cross-validation (Section S1.6). If the effective sample size is smaller than the nominal sample size because of autocorrelation, we will tend to include too small a proportion of neighbours, resulting in undersmoothed parameter estimates. This would increase the tendency for reefs in different regions of the simplex to move in different directions, and would therefore increase rather than decrease the chance of concluding that there were alternative stable states.

2. Each sequence of observations on a single reef at consecutive one-year intervals is a realization of the same stochastic process. We are not able to model heterogeneity among reefs, because we have a relatively small number of observations on any single reef.

3. The future states of reefs are conditionally independent of past states, given the current state. This is the Markovian assumption, which we make because it greatly simplifies the statistical modelling. In reality, it is unlikely to be precisely true. For example, the relative abundances of coral species on a reef that has had high coral cover for a long time may differ from those on a reef that has had high coral cover for a short time, potentially affecting dynamics (Connell, 1997). Such variation in community composition and evenness could influence the relative abundance of disturbance-sensitive species (Rogers, 1993). Additionally, larger colonies and massive boulder species are also more likely on reefs that have had high coral cover for a long time, and vulnerability to competition, predation, and storm damage varies with colony size (Tanner et al., 1996; Madin and Connolly, 2006).

4. The stochastic process generating the sequences of observations is homogeneous in time. Thus, information about past dynamics tells us directly about future dynamics. Our model is stochastic, so that it includes the effects of chance processes such as storms, but we assume that the statistical properties of these chance processes have remained constant.

5. We can make predictions about the fate of a reef in a given state from the fates of reefs with similar states. The more similar two states are, the more useful they are likely to be as predictors of each other's fates. We can therefore use local linear estimation to make predictions about the underlying nonlinear time series. Related methods are reviewed by Fan and Gijbels (1996, chapter 6).

6. The measurement error in reef state is small relative to the inherent stochasticity in fate for a given state. Thus we treat the true state at a given time as known. In reality, it is likely that measurement error is smaller than inherent stochasticity, but still substantial. However, including both kinds of error would require much more complex analyses (Doucet et al., 2001). Instead, in Sections S1.13 and S2.8, we check the robustness of our methods to measurement error by simulation.

S1.3 Half-taxi distance

Measures of distance on the simplex should ideally take account of the geometry of the simplex space, in which components are non-negative and sum to 1 (Aitchison, 1992; Aitchison and Egozcue, 2005).

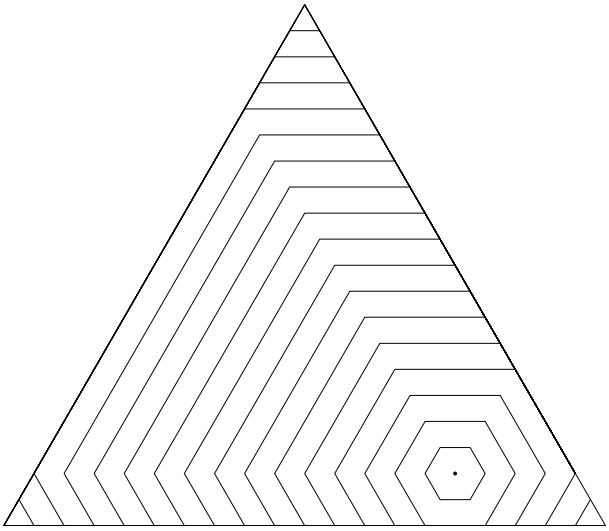

Figure S1: Contours of the half-taxi distance on the 2-simplex at intervals of 0.05 around the point $(0.7, 0.1, 0.2)$.

A commonly used distance for compositional data is *Aitchison distance* (Aitchison, 1992)

$$d_A(\mathbf{x}, \mathbf{z}) = \sqrt{\sum_{j=1}^{3} \left(\log(x_j/m_g(\mathbf{x})) - \log(z_j/m_g(\mathbf{z}))\right)^2},$$

where $\mathbf{x}$ and $\mathbf{z}$ are two compositions with components x_j and z_j and $m_g(\mathbf{x})$ denotes the geometric mean, i.e., $m_g(\mathbf{x}) = (x_1 x_2 x_3)^{1/3}$. However, contours of the Aitchison distance become very closely spaced as we approach the edges of the simplex, and if any component has zero value, the Aitchison distance is undefined (Miller, 2002). This is undesirable for our data. We have many points close to the edges of the simplex, and intuitively, we want those with small absolute differences in all components to be separated by a small distance.

An alternative distance measure whose properties seem more appropriate for our data is the *half-taxi distance* (Miller, 2002)

$$d_T(\mathbf{x}, \mathbf{z}) = \frac{1}{2}\left(|x_1 - z_1| + |x_2 - z_2| + |x_3 - z_3|\right).$$

Contours of the half-taxi distance remain evenly spaced as we approach the edges of the simplex (Figure S1). We will therefore use the half-taxi distance to measure differences in reef state.

S1.4 Dirichlet transition kernel

The Dirichlet distribution is one of the simplest distributions on the 2-simplex, requiring only 3 parameters. We therefore think it is appropriate for a first attempt at modelling the transition kernel (main text, Equation 2). The Dirichlet density at a point $\mathbf{q} = [q_1, q_2, q_3]$ is

$$\text{Dirichlet}(\mathbf{q}; \boldsymbol{\alpha}) = \frac{1}{Z(\boldsymbol{\alpha})} \prod_{j=1}^{3} q_j^{\alpha_j - 1},$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3] > 0$ are the parameters,

$$Z(\boldsymbol{\alpha}) = \frac{\prod_{j=1}^{3} \Gamma(\alpha_j)}{\Gamma(\sum_{j=1}^{3} \alpha_j)}$$

and $\Gamma()$ is the gamma function. Thus, we use the transition kernel

$$k(\mathbf{y}|\mathbf{x}) \sim \text{Dirichlet}(\boldsymbol{\alpha}(\mathbf{x})), \tag{S1}$$

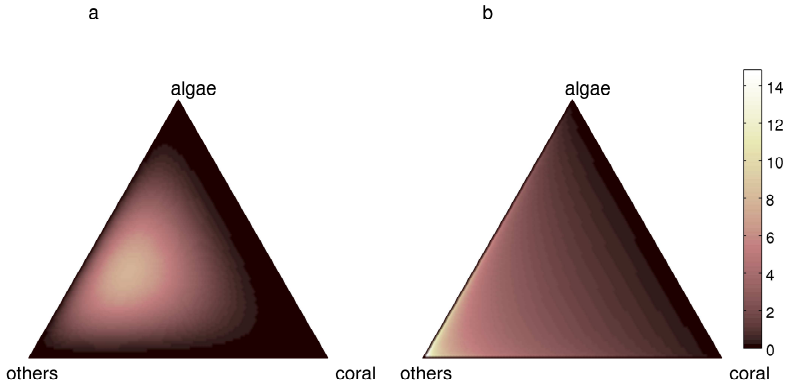

Figure S2: (a) The Dirichlet$(2, 3, 4)$ distribution. (b) The Dirichlet$(0.75, 0.9, 2)$ distribution.

where the values of the parameters $\boldsymbol{\alpha}(\mathbf{x})$ depend on the state $\mathbf{x}$.

To calculate the centre of a Dirichlet distribution (Aitchison, 1989), which we use as a summary of location, we need the marginal geometric means of each component. The marginal distribution of the ith component in a Dirichlet distribution is a Beta(ν, ω) distribution, with $\nu = \alpha_i$ and $\omega = \sum_{j \neq i} \alpha_j$. This has geometric mean $\psi(\nu) - \psi(\nu + \omega)$, where $\psi()$ is the digamma function.

The Dirichlet distribution can have at most one interior mode (Figure S2a), or can have modes on the edges of the simplex (Figure S2b). It is unable to describe patterns such as banana-shaped groups of observations on the interior of the simplex (Aitchison, 1986, Figure 1.9 and Section 3.4). Thus, there is scope in future for using more flexible distributions (Aitchison, 1986, chapter 6), but we should proceed cautiously because of the extra parameters these distributions require.

S1.5 Local linear estimation

We do not have multiple observations of the fates of reefs with current state $\mathbf{x}$ from which to estimate the parameters of Equation S1. However, we do have large numbers of observations, of which some come from reefs with states that are in some sense close to $\mathbf{x}$. We will use the fates of these nearby points to estimate $k(\mathbf{y}|\mathbf{x})$.

Let $\mathbf{p}$ denote a point on the simplex at which we want to estimate the update distribution. We have a set of observations $\mathcal{X} = \{\mathbf{x}_i\}$, $i = 1, 2, \ldots, n$, around $\mathbf{p}$, and the fates $\mathcal{Y} = \{\mathbf{y}_i\}$ of these observations one year later. The point $\mathbf{p}$ may or may not be in the set $\mathcal{X}$. For example, $\mathbf{p}$ will not be in $\mathcal{X}$ if it is an arbitrary point at which we do not have an observation.

We want to allow the update distribution to depend on $\mathbf{x}$, but we do not want to specify the form of this dependency in advance. Instead, we will use a local estimate of $\boldsymbol{\alpha}$, which allows the parameters to change over the simplex, and gives most weight to observations close to $\mathbf{p}$. In general terms, we want functions $g_j()$ such that $\alpha_j = g_j(\mathbf{p})$. Since we require that $\alpha_j > 0$, we also need a link function which ensures that this condition is always met. An obvious choice is the log link function so that $\alpha_j = \exp(\eta_j(\mathbf{p}))$, where $\eta_j()$ is a smooth (i.e. having continuous second derivative) function which can take any real value. The elements of $\mathbf{p}$ sum to 1, hence the last element is redundant ($p_3 = 1 - p_1 - p_2$), and we can write $\alpha_j = \exp(\eta_j(p_1, p_2))$. We do not assume any particular parametric form of the function η_j. We will assume, however, that this function is smooth, so that for starting points which are close, the parameters $\boldsymbol{\alpha}$ are also similar.

Taylor's theorem states that a smooth function can be approximated locally by a polynomial with coefficients depending on the derivatives of the function. Thus, if a function $g(z_1, z_2)$ has continuous second derivatives, then it can be approximated by a linear function in the neighbourhood of a point

(z_1, z_2) by

$$g(z_1+\delta_1, z_2+\delta_2) \;=\; g(z_1,z_2) + \frac{\partial g}{\partial z_1}\Big|_{z_1,z_2}\delta_1 + \frac{\partial g}{\partial z_2}\Big|_{z_1,z_2}\delta_2 + R(z_1,z_2),$$

where the error term $R(z_1, z_2)$ is proportional to a quadratic form in (δ_1, δ_2).

Since η_j is assumed to be a smooth function, it can be approximated locally by a linear function

$$\begin{aligned}\eta_j(x_{i1}, x_{i2}) &\approx \eta_j(p_1,p_2) + \frac{\partial \eta_j}{\partial x_{i1}}\Big|_{p_1,p_2}(x_{i1}-p_1) + \frac{\partial \eta_j}{\partial x_{i2}}\Big|_{p_1,p_2}(x_{i2}-p_2)\\ &= \beta_{0j\mathbf{p}} + \beta_{1j\mathbf{p}}(x_{i1}-p_1) + \beta_{2j\mathbf{p}}(x_{i2}-p_2) = \eta_j(\mathbf{p}, i). \qquad \text{(S2)}\end{aligned}$$

Obviously, the closer the point (x_{i1}, x_{i2}) is to (p_1, p_2) the better is the approximation.

Then the unknown parameters $\boldsymbol{\beta}_\mathbf{p} = \{\beta_{0j\mathbf{p}}, \beta_{1j\mathbf{p}}, \beta_{2j\mathbf{p}}\}, j = 1, 2, 3$ are the intercept and slopes with respect to p_1 and p_2 of a linear approximation to $\log(\alpha_j)$ close to a point $\mathbf{p}$. These parameters are estimated by maximizing the local log-likelihood defined as follows. Denote $\alpha_j(\mathbf{p}, i) = \exp(\eta_j(\mathbf{p}, i))$, then for a pair $(\mathbf{x}_i, \mathbf{y}_i)$, the log likelihood of $\mathbf{y}_i$ conditional on $\mathbf{x}_i$, for parameter estimates $\boldsymbol{\alpha}(\mathbf{p}, i) = [\alpha_1(\mathbf{p}, i), \alpha_2(\mathbf{p}, i), \alpha_3(\mathbf{p}, i)]$ is

$$l_{i\mathbf{p}} = -\log(Z(\boldsymbol{\alpha}(\mathbf{p}, i))) + \sum_{j=1}^{3}(\alpha_j(\mathbf{p}, i) - 1)\log y_{ij}, \qquad \text{(S3)}$$

where y_{ij} is the jth component of $\mathbf{y}_i$.

Notice that if any $y_{ij} = 0$, then l_{ip} is undefined. In our data, there was only one such observation and hence we used an ad hoc approach of offsetting this value by a small number ($\varepsilon = 10^{-12}$) and subtracting $\varepsilon/2$ from the other two elements of $\mathbf{y}_i$ (to ensure that $\sum_j y_{ij} = 1$). Other values of ε were also considered and the results were similar. If there were more observations with zeros in the data, one should consider a more systematic approach (Bacon-Shone, 2003).

Under the assumptions that reefs are independent, the process of change in reef state is homogeneous and Markovian, and that the more similar two states are, the more information the fate of a reef in one of those states gives us about the distribution of fates for a reef in the other state, we can obtain a local estimate of the parameters $\boldsymbol{\beta}_\mathbf{p}$ by maximizing a weighted sum of the $l_{i\mathbf{p}}$ over all the observations in $\mathcal{Y}$:

$$l_\mathbf{p} = \sum_{i=1}^{n} v_{i\mathbf{p}} l_{i\mathbf{p}} \qquad \text{(S4)}$$

where $v_{i\mathbf{p}} > 0$ is the weight associated with point $\mathbf{x}_i$ as a source of information about the fate of point $\mathbf{p}$. Since the closer the point (x_{i1}, x_{i2}) is to (p_1, p_2), the better the approximation in Equation S2, the weight $v_{i\mathbf{p}} > 0$ is an decreasing function of the distance between these points.

The weights $v_{i\mathbf{p}}$ are defined as a function of the distance between points $\mathbf{x}$ and $\mathbf{p}$

$$v_{i\mathbf{p}} = v(\mathbf{x}_i, \mathbf{p}) = K\left(\frac{d_T(\mathbf{x}_i, \mathbf{p})}{h}\right),$$

where K is a kernel function and h is a bandwidth. The kernel is usually a symmetric non-negative function. It is often chosen so that it takes positive values in a finite interval, so that the bandwidth governs the support of the kernel. It is convenient to assume that the support of the kernel is $[-1, 1]$, then only the points $\mathbf{x}_i$ for which $d_T(\mathbf{x}_i, \mathbf{p}) < h$ influence the estimates of the parameters $\boldsymbol{\beta}_\mathbf{p}$ at point $\mathbf{p}$. Here the Epanechnikov kernel $K(x) = 0.75(1 - x^2)$ (Figure S3) was used (Fan and Gijbels, 1996, p. 15).

We maximized $l_\mathbf{p}$ using a BFGS quasi-Newton algorithm (Nocedal and Wright, 1999, chapter 6) with a cubic line search procedure, as implemented in the Matlab function `fminunc` (Matlab Optimization Toolbox Version 4.3, The Mathworks Inc., Natick, MA).

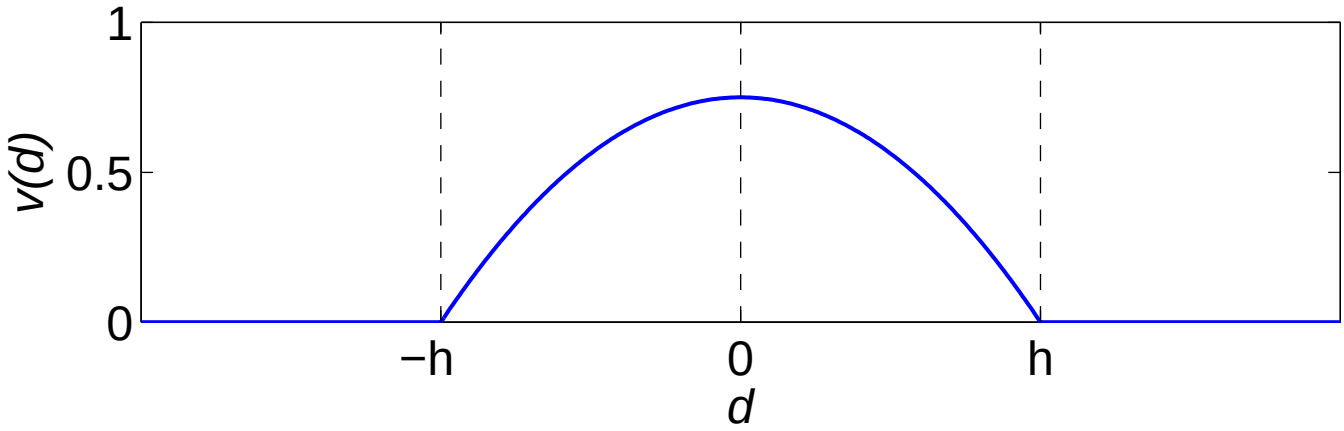

Figure S3: The Epanechnikov kernel $v(d)$ with bandwidth h, as a function of distance d.

S1.6 Choice of bandwidth

The bandwidth plays a crucial role in local linear estimation. In the limit of infinite bandwidth, all the points receive equal weights in the local log-likelihood, which then becomes equivalent to the global log-likelihood. For a very small bandwidth, the estimate is based on very few points in the neighbourhood of $\mathbf{p}$ and hence might be very variable and unstable. In particular, if h is so small that there are no observation in the neighbourhood of $\mathbf{p}$, then there is no information on which to base the local estimate at this point.

If the observations were uniformly distributed on the simplex, it might be appropriate to use the same value of the bandwidth everywhere, as then the number of observations on which parameter estimates were based would be similar for all points. However, in our data, there are some regions where observations are very dense and others where they are very sparse. We therefore allow the bandwidth to vary so that the number of points used in the estimation is the same at each point $\mathbf{p}$. More precisely, if the number of points which are to be included in the local neighbourhood is n_0, then the bandwidth for this point is defined as $h_{\mathbf{p},n_0} = d_{n_0}$, where d_{n_0} is the distance between the (n_0+1)th nearest neighbour and $\mathbf{p}$. Notice that this means that the weight in the local log-likelihood for the(n_0+1)th neighbour is zero.

The optimal bandwidth yields the estimated densities closest to the true densities, but since the true densities are unknown this cannot be found directly. Instead, several automatic selection criteria exist for choosing the bandwidth, of which we used the cross-validated deviance (Hastie and Tibshirani, 1990),

$$\mathrm{Dev}_{\mathrm{CV}} = \frac{2}{n}\sum_{i=1}^{n}\left(\log(\mathrm{Dirichlet}(\mathbf{y}_i;\hat{\boldsymbol{\alpha}}_{\max})) - \log(\mathrm{Dirichlet}(\mathbf{y}_i;\hat{\boldsymbol{\alpha}}_{-i}))\right)$$

where $\hat{\boldsymbol{\alpha}}_{-i}$ are estimates of the parameters at $\mathbf{x}_i$ based on the sample with ith observation excluded and $\hat{\boldsymbol{\alpha}}_{\max}$ are the estimates for the maximal model (the model for which the degrees of freedom are equal to the number of observations), that do not depend on the bandwidth h (Hastie and Tibshirani, 1990). The optimal bandwidth is the one that minimizes $\mathrm{Dev}_{\mathrm{CV}}$. In practice, it is easier to work with the proportions of nearest neighbours that should be included in the estimation than with their number. We optimized this proportion using a golden section search (Press et al., 1992, section 10.1) as implemented in the Matlab function `fminbnd` (Matlab Optimization Toolbox Version 4.3, The Mathworks, Inc., Natick, MA). The estimated proportion for Caribbean data was 0.664, for Kenyan data 0.627, and for the Great Barrier Reef it was 0.288.

S1.7 The transition equation

Given a method for estimating the transition kernel between any two points on the simplex, and the density f_t over the simplex at time t, we can express the density f_{t+1} at any point $\mathbf{q}$ at time $t+1$ as

an integral equation:

$$f_{t+1}(\mathbf{q}) = \int k(\mathbf{q}|\mathbf{p}) f_t(\mathbf{p})\, d\mathbf{p} \tag{S5}$$

where the integral is over the entire simplex. Equation S5 says that at time $t+1$, the density at any point is determined by the integral of transitions from all other points on the simplex, weighted by the densities at those points at time t. Equation S5 is a linear homogeneous Fredholm integral equation of the first kind (Baker, 1977, p. 3).

S1.8 Solving the transition equation

It is not usually possible to solve Equation S5 analytically. Of the numerical methods available, quadrature methods (Baker, 1977, pp. 169-170) are among the simplest and most popular. In these methods, we approximate the integral in Equation S5 by a weighted sum

$$f_{t+1}(\mathbf{q}) \approx \tilde{f}_{t+1}(\mathbf{q}) = \sum_{j=1}^{N} w_j k(\mathbf{q}|\mathbf{p}_j) f_t(\mathbf{p}_j) \tag{S6}$$

where the N Gauss points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N$ and their corresponding weights $w_1, w_2, \ldots, w_N$ are chosen by Gaussian quadrature on the simplex (von Winckel, 2005).

Equation S6 lets us solve for one time step. To understand the long-term behaviour of the model, we want to find eigenfunctions $f(\mathbf{q})$ of Equation S5 such that

$$\kappa f(\mathbf{q}) = \int k(\mathbf{q}|\mathbf{p}) f(\mathbf{p})\, d\mathbf{p}$$

for eigenvalues κ. In particular, because the integral of any probability density function will always be 1, we know that the largest eigenvalue will be 1, and its associated eigenfunction will be the equilibrium (stationary) distribution of the model. We know that the model will converge to this stationary distribution, provided we extend by continuity to the closed simplex, because we then have a continuous, power-positive kernel on a compact and bounded metric space (Ellner and Rees, 2006, Appendix C).

Let $\mathbf{K} = \{k(\mathbf{p}_i|\mathbf{p}_j)\}$, $\mathbf{D} = \mathrm{diag}(w_1, w_2, \ldots, w_N)$, and $\tilde{\mathbf{f}} = [\tilde{f}(\mathbf{p}_1), \tilde{f}(\mathbf{p}_2), \ldots, \tilde{f}(\mathbf{p}_N)]^T$. Then we can find the eigenvalues $\tilde{\kappa}$ and eigenvectors $\tilde{\mathbf{f}}$ of the matrix equation

$$\mathbf{K}\mathbf{D}\tilde{\mathbf{f}} = \tilde{\kappa}\tilde{\mathbf{f}}.$$

The sth eigenvector is an approximation to the corresponding eigenfunction at the points $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N$. We can then evaluate the approximate eigenfunctions at any other point $\mathbf{x}$ using Nyström's method (Baker, 1977, p. 170):

$$\tilde{f}_s(\mathbf{x}) = \frac{1}{\tilde{\kappa}_s} \sum_{j=1}^{N} w_j k(\mathbf{x}|\mathbf{p}_j) \tilde{f}_s(\mathbf{p}_j). \tag{S7}$$

In particular, the solution to Equation S7 for the value of s for which $\tilde{\kappa}_s \approx 1$ approximates the stationary distribution of the model. The approximation gets more accurate as the number of Gauss points increases. We chose 1600 points, which gave largest approximate eigenvalues of 0.9998 for the Caribbean, 0.9983 for Kenya, and 0.9966 for the Great Barrier Reef.

S1.9 Summary statistics

We use similar quadrature approximations to calculate summary statistics (centre and square root of generalized variance) for the estimated stationary distribution. Standard statistics such as the arithmetic mean and untransformed covariance matrix are difficult to interpret for compositional data. An appropriate measure of the centre for a compositional distribution with density function $f(\mathbf{p})$ is

$$\mathbf{c} = \mathcal{C}\left[\left\{\exp \int f(p_i) \log(p_i)\, \mathrm{d}p_i\right\}\right],$$

(Aitchison, 1989), where $\mathcal{C}(\mathbf{w}) = \mathbf{w}/(\sum_i w_i)$ (the closure operator), and $f(p_i)$ is the marginal density of the ith part of the composition. We can approximate this by

$$\mathbf{c} \approx \mathcal{C}\left[\left\{\exp\sum_{j=1}^{N} w_j \tilde{f}_s(p_{ij})\log(p_{ij})\right\}\right], \tag{S8}$$

where p_{ij} is the value of the ith part of the composition at the jth Gauss point.

As a measure of spread, we use the square root of the generalized variance (determinant) of the logratio covariance matrix, $\sqrt{|\mathbf{\Sigma}|}$. For a D-dimensional composition x, the elements of $\mathbf{\Sigma}$ are $\sigma_{ij} = \mathrm{cov}\left(\log(x_i/x_D), \log(x_j/x_D)\right), i, j = 1 \ldots D-1$ (Aitchison, 1986, pp. 76-78). We use a quadrature approximation similar to Equation S8 to obtain $\mathbf{\Sigma}$.

It would be possible to do formal tests of hypotheses about differences in stationary distributions between regions. We have not done so, because there is strong evidence for such differences between the Caribbean and Great Barrier Reef from a much simpler model with discretized reef states (P. Lowe et al., in preparation). For the model described here, the necessary parametric bootstrap tests are very computationally intensive, and the null hypothesis of no difference in stationary distributions seems unlikely to be true.

S1.10 Residuals

Residuals are an important tool in evaluating the assumptions of the model and its performance. Gueorguieva et al. (2008) describe several kinds of residuals for Dirichlet-distributed data. Because the Dirichlet distribution is multivariate, we can consider both univariate residuals (for each component of the distribution separately) and multivariate residuals (that summarize the difference between observed and expected values over all components).

The univariate residual we use is the standardized residual for each component j of observation i:

$$r_{ij} = \frac{y_{ij} - E(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i))}{\sqrt{\mathrm{Var}(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i))}}$$

where $\hat{\boldsymbol{\alpha}}(\mathbf{x}_i)$ is a vector of estimated Dirichlet parameters for observation i, obtained as described in Section S1.3, and Y_{ij} is the random variable distributed as the jth component of the fate of observation i. The expected value of the jth component, given the estimated parameters, is $E(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i)) = \hat{\alpha}_j(\mathbf{x}_i)/\sum_m \hat{\alpha}_m(\mathbf{x}_i)$, and $\mathrm{Var}(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i)) = [E(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i))(1 - E(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i)))]/[1 + \sum_m \hat{\alpha}_m(\mathbf{x}_i)]$ is the variance of the jth component of a Dirichlet distribution with the estimated parameters. Standardization is necessary because the raw residuals $y_{ij} - E(Y_{ij}; \hat{\boldsymbol{\alpha}}(\mathbf{x}_i))$ have different variances for different components and Dirichlet distributions, and are therefore hard to compare (Gueorguieva et al., 2008).

Positive standardized univariate residuals indicate that the observed value of a given component was larger than expected, and vice versa. If the model is appropriate, the standardized univariate residuals should have sample means close to 0 and sample variances close to 1.

Given the standardized univariate residuals, we can then form the composite residual $C_i = \sum_m r_{im}^2$, which measures the overall departure of the observed from the expected value, giving equal weight to each component (Gueorguieva et al., 2008). Composite residuals are always non-negative, with larger values indicating larger differences between observed and expected values. The other, more complicated residuals discussed by Gueorguieva et al. (2008) were highly correlated with the ones used here, for the data they considered.

If the model is performing well, there should be no relationship between residuals and state. We can investigate this by plotting the standardized univariate residuals against each component of the state at time t. To reveal any trends, we used a LOWESS (locally-weighted least squares) smoother (Cleveland, 1979) implemented in the `smooth` function in the Matlab Curve-Fitting Toolbox version 2.1 (The Mathworks, Inc., Natick, MA), with a first degree polynomial model and span chosen by eye to give sufficient smoothing.

We also use the residuals to check some of the assumptions listed in the main text:

- If each reef is a realization of the same process, there should be no systematic differences in residuals among reefs.

- We have not developed a direct test of the Markovian assumption that future states are conditionally independent of past states, given the current state. However, examining scatter plots of the residuals (at time $t+1$) against the state of the reef at time $t-1$ (the year before the observation whose fate we are predicting), with a LOWESS smoother as above, may give us some insights into possible dependencies on past state.

- If the process is homogeneous in time, the distribution of residuals should be the same for all observation times.

S1.11 Estimates of uncertainty

In order to estimate the uncertainty associated with the equilibrium distribution, we performed a leave-one-out jackknife for each region. Within each region, we used the methods described above to obtain an estimate $f_{(i)}(\mathbf{p})$ of the equilibrium density at any point $\mathbf{p}$ without the ith pair of state and fate ($i = 1, 2, \ldots, n$, where n is the number of fate and state pairs). The jackknife standard error of the equilibrium density at $\mathbf{p}$ is then

$$\hat{\text{s.e.}}\,[f(\mathbf{p})] = \left\{ \frac{n-1}{n} \sum_{i=1}^{n} \left[f_{(i)}(\mathbf{p}) - f_{(.)}(\mathbf{p}) \right]^2 \right\}^{1/2} ,$$

where

$$f_{(.)}(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^{n} f_{(i)}(\mathbf{p}).$$

This method is conservative in the sense that it is likely to overestimate the standard error (Efron and Stein, 1981). The bootstrap does not perform well in this case, because the presence of exact duplicate observations adversely affects the performance of leave-one-out cross-validation.

S1.12 Estimation from a model with alternative stable states

To investigate whether we can detect alternative stable states when we know they exist, we simulate data under a stochastic differential equation version of the simple model described in Mumby et al. (2007), and analyze these data using the same methods that we applied to the real data.

The basic deterministic model has three state variables, C, M, and T, the proportional cover of corals, macroalgae, and algal turfs, respectively ($C + M + T = 1$). We treat C, M, and T as being analogous to our coral, algae, and other components, respectively. The system of differential equations studied by Mumby et al. (2007) is

$$\begin{aligned} \frac{dC}{dt} &= rTC - \mu C - aMC \quad , \\ \frac{dM}{dt} &= aMC - \frac{gM}{M+T} + \gamma MT \quad . \end{aligned} \tag{S9}$$

Here, r is the recruitment rate of corals over algal turfs, μ is the natural mortality rate of corals, a is the rate of coral overgrowth by macroalgae, g is the rate of grazing on macroalgae, and γ is the rate at which macroalgae overgrow algal turfs. This deterministic model has alternative stable states (Figure S4) for some parameter values (Mumby et al., 2007).

To produce an analogous model with environmental stochasticity, we first transform from a composition $[C, M, T]$ on the 2-simplex into a pair of new state variables A and B on the real plane using the additive logratio transformation (Aitchison, 1986, p. 113):

$$\begin{aligned} A(t) &= \log \frac{C(t)}{T(t)} \quad , \\ B(t) &= \log \frac{M(t)}{T(t)} \quad . \end{aligned}$$

Working on the real plane will allow us to introduce stochasticity in the form of standard Wiener processes (Higham, 2001, section 2), which is simpler than trying to define a stochastic process directly

on the simplex. The deterministic dynamics of A, B are given by

$$\begin{aligned} \frac{dA}{dt} &= \frac{dC}{dt}\frac{\partial A}{\partial C} + \frac{dM}{dt}\frac{\partial A}{\partial M} &= f_A(C, M, T) \quad , \\ \frac{dB}{dt} &= \frac{dC}{dt}\frac{\partial B}{\partial C} + \frac{dM}{dt}\frac{\partial B}{\partial M} &= f_B(C, M, T) \quad , \end{aligned} \tag{S10}$$

where $\partial A/\partial C = (1 - M)/CT$, $\partial A/\partial M = 1/T$, $\partial B/\partial C = 1/T$, $\partial B/\partial M = (1 - C)/MT$. We then define a system of stochastic differential equations analogous to Equation S10:

$$\begin{aligned} dA(t) &= f_A(C, M, T)dt + \psi_A dW_A t \quad , \\ dB(t) &= f_B(C, M, T)dt + \psi_B dW_B t \quad . \end{aligned} \tag{S11}$$

The stochastic terms W_A and W_B are independent standard Wiener processes. The noise coefficients ψ_A and ψ_B determine the amount of noise in the system. We can simulate Equation S11 using the Euler-Maruyama algorithm (Higham, 2001, section 4). Finally, we can back-transform onto the simplex:

$$\begin{aligned} C(t) &= \frac{e^{A(t)}}{e^{A(t)} + e^{B(t)} + 1} \quad , \\ M(t) &= \frac{e^{B(t)}}{e^{A(t)} + e^{B(t)} + 1} \quad , \\ T(t) &= 1 - C(t) - M(t) \quad . \end{aligned}$$

Figure S5 is a stochastic analogue of Figure S4. The system has two basins of attraction, with trajectories approaching one or the other along a diagonal band.

The model defined by Equation S11 is not the only way of producing a stochastic process on the simplex whose deterministic skeleton has alternative stable states (Blackwood et al., 2011). However, it is one possible way of generating suitable test data. We simulated time series from Equation S11, using the same initial values as the real Caribbean data, and sampled at regular intervals to produce discrete time series with the same number of observations as the real data. The noise coefficients ψ_A and ψ_B were set to 0.4, and one year was set to five units of time, in order to generate time series with amounts of variability and annual change that looked similar to those in the real data. Other parameters were the same as those used to generate Figure 3a in Mumby et al. (2007)(P. Mumby, personal communication). We ran 100 replicate simulations of the stochastic model with parameters as in Figure S5. We then applied the methods we used for the real data to each replicate simulation. We scored each stationary distribution as unimodal, bimodal, multimodal, or other by eye.

S1.13 Measurement error

In addition, we investigated the effects of measurement error on analyses based on the stochastic model described above. We searched for papers containing information on sampling variability in coral cover on electronic databases including Web of Knowledge, Google Scholar, and ReefBase. We restricted our searches to coral cover because preliminary investigation suggested there were few data on sampling variability in other groups such as macroalgae. We therefore assumed that measurement error in other groups behaved in a similar way. We also concentrated on sampling methods in which the state of the substrate is recorded at discrete points. Such methods include the Line Point Intercept (LPI) method, photoquadrats with point sampling, and video transects with point sampling (Aronson et al., 1994; Leujak and Ormond, 2007). Simulations suggest that the qualitative pattern of sampling error is similar for other methods (Nadon and Stirling, 2006). A meta-analysis also showed that measurement error in the rate of change of coral cover was fairly consistent between methods, although there was more variability in the estimated rate of change of macroalgal cover (Côté et al., 2005).

We extracted the mean coral cover and its coefficient of variation among sampling units (transects or quadrats) from each paper (Table S1) by measuring histograms.

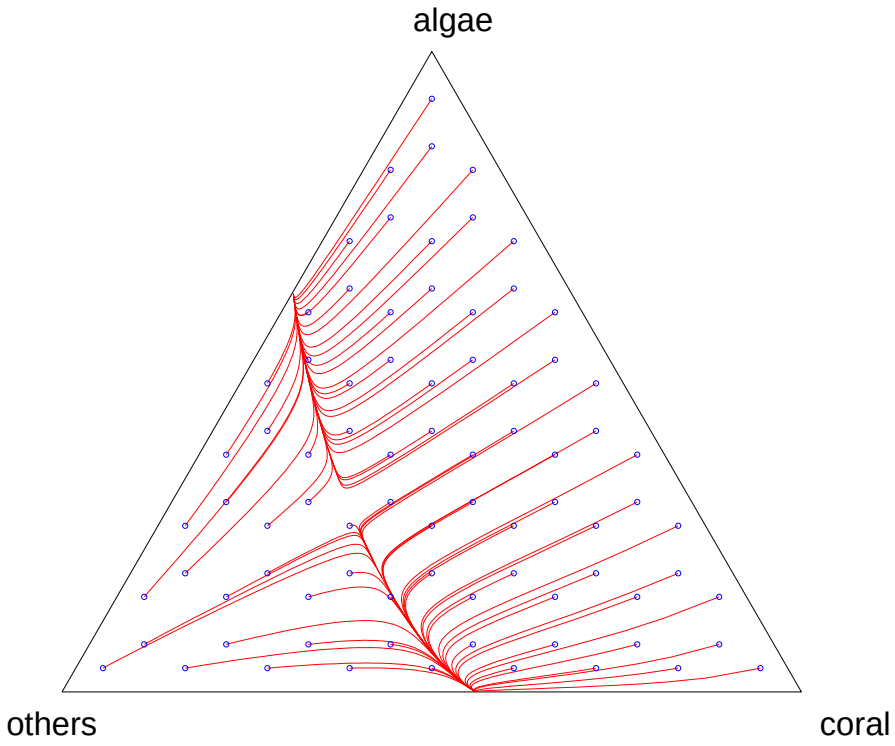

Figure S4: Numerical solutions (red lines) of the deterministic model specified by Equation S9. Blue circles are initial values. Redrawn from Mumby et al. (2007). Parameter values: $a = 0.1$, $g = 0.3$, $\gamma = 0.8$, $\mu = 0.44$, $r = 1$.

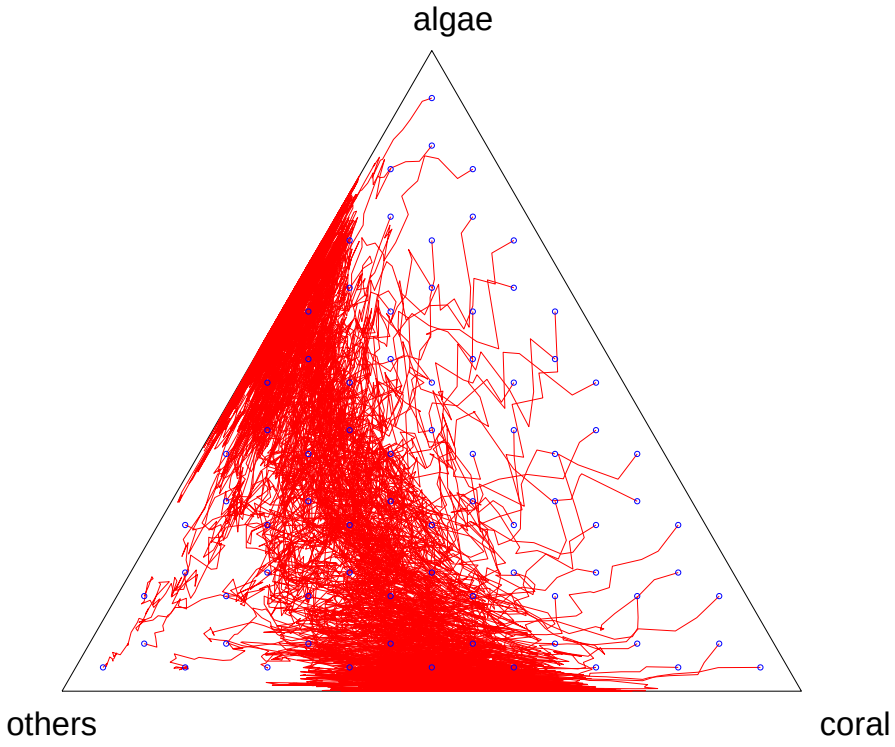

Figure S5: One simulated realization (red line) of the stochastic model specified by Equation S11 for each member of a grid of initial values (blue circles). Parameter values: $\psi_A = \psi_B = 0.4$. Other parameters as in Figure S4.

Table S1: Literature data on mean and coefficient of variation (CV) of coral cover, from video transects, photoquadrats, and Line Point Intercept (LPI). Points per unit is the number of points at which the state of the substrate is recorded in each sampling unit (quadrat or transect).

Study	Mean	CV	Method	Points per unit	Reef
Aronson et al. (1994)[1]	0.03	0.58	Video	500	Discovery Bay, Jamaica
	0.16	0.24	Video	500	Carrie Bow Cay, Belize
Edmunds and Bruno (1996)[2]	0.03	1.00	Photoquadrat	200	Mooring One, Jamaica
	0.06	0.25	LPI	100	Blowing Rocks, Jamaica
	0.24	0.75	Photoquadrat	200	Dairy Bull, Jamaica
	0.24	0.69	Photoquadrat	200	Silver Spray, Jamaica
	0.33	0.73	Photoquadrat	200	Yawzi Point, St. John
	0.05	1.00	Photoquadrat	200	White Point, St. John
	0.09	0.67	Photoquadrat	200	West Tektite, St. John
Harding et al. (2000)[3]	0.27	0.09	LPI	80	Pulau Balak, Sabah
Leujak and Ormond (2007)[4]	0.46	0.43	LPI	50	Ras Um Sidd, Egypt
	0.38	0.49	Photoquadrat	100	Ras Um Sidd, Egypt
	0.39	0.48	Video	204	Ras Um Sidd, Egypt
Murdoch and Aronson (1999)[5]	0.22	0.16	Video	500	Bir Key Reef, Florida
	0.19	0.23	Video	500	Pulaski Shoal, Florida
	0.08	0.34	Video	500	Cosgrove Shoal, Florida
	0.05	0.18	Video	500	Sand Key Reef, Florida
	0.07	0.38	Video	500	Western Sambo Reef, Florida
	0.14	0.30	Video	500	American Shoal, Florida
	0.11	0.23	Video	500	Looe Key Reef, Florida
	0.10	0.26	Video	500	No Name Reef, Florida
	0.02	0.70	Video	500	Sombrero Key Reef, Florida
	0.11	0.32	Video	500	Love Reef, Florida
	0.10	0.33	Video	500	Tennessee Reef, Florida
	0.16	0.26	Video	500	Alligator Reef, Florida
	0.08	0.41	Video	500	Pickles Reef, Florida
	0.05	0.32	Video	500	The Elbow, Florida
	0.02	0.45	Video	500	Carysfort Reef, Florida
	0.09	0.36	Video	500	Ajax Reef, Florida
Nadon and Stirling (2006)[6]	0.01	1.30	LPI	20	North Bellairs, Barbados
	0.02	0.95	LPI	20	North Bellairs, Barbados
	0.10	0.45	LPI	20	North Bellairs, Barbados

[1] their Figure 1, 10 points per frame
[2] their Figure 2
[3] their Figure 2, coefficient of variation calculated from difference between staff and volunteers in week 4
[4] their Figure 2a, 95% confidence interval assumed to be 1.96 standard deviations
[5] their Figure 5
[6] their Figure 2

If we assume that each sampled point is independent of all the others, and that within a survey, the probability of being covered by each possible category is the same for all m points, then the obvious model for measurement error is the multinomial. Denote by $\mathbf{z}(t)$ the number of sampled points in each state at time t. Then

$$\mathbf{z}(t) \sim \text{multinomial}(m, \mathbf{x}(t)), \tag{S12}$$

where $\mathbf{x}(t)$ is the true state at time t. The maximum likelihood estimate of the true state probabilities $\mathbf{x}(t)$ is

$$\hat{\mathbf{x}}(t) = \frac{1}{m}\mathbf{z}(t), \tag{S13}$$

and the coefficient of variation for the ith state is

$$CV\left(\frac{z_i}{m}\right) = \frac{1}{\sqrt{m}}\sqrt{\frac{1-x_i}{x_i}}. \tag{S14}$$

Where there are multiple data points from the same LPI ($m = 20$) or video ($m = 500$) study, the multinomial approximates the qualitative relationship between CV and mean (Figure S6). However, the CV from photoquadrats ($m = 200$) is consistently well above that predicted by the multinomial model. This may be because the points sampled within a photoquadrat are close together, and are therefore affected by small-scale spatial aggregation. We used the multinomial as a model for measurement error, with the caveat that the assumption of independent sampling points may result in an underestimate of variability.

We repeated the simulations from the stochastic model with parameters as in Figure S5 , sampling between 20 and 5000 points at each observation time under the multinomial model (Equation S12), and using Equation S13 as an estimator of state. The lower limit of 20 points is much lower effort than would typically be used in practice, although it has been used to investigate the effects of measurement error (Nadon and Stirling, 2006). The Reef Check programme uses a relatively low sampling effort of 160 points along 4×20 m transects (Leujak and Ormond, 2007). 5000 points is the number recommended by Aronson et al. (1994), corresponding to 10 points on each of 50 video frames, on each of 10×25 m transects. Most of our data are likely to have sampling effort corresponding to somewhere between 160 and 5000 points.

We analyzed the simulated data in the same way as the real data, and scored each stationary distribution as unimodal, bimodal, trimodal, or other by eye. In addition, we reanalyzed the Caribbean and Great Barrier Reef regions without the Reef Check data, which are likely to lie at the lower end of the sampling effort distribution. Reef Check contributed 68 out of 100 pairs of observations for the Caribbean, but only 10 out of 374 for the Great Barrier Reef.

S1.14 Implementation

We implemented the methods described here in Matlab version 7.9 (The Mathworks, Inc., Natick, MA), with the Matlab Curve Fitting Toolbox Version 2.1, Optimization Toolbox version 4.3, and Statistics Toolbox version 7.2. The code is available from:

`http://www.liv.ac.uk/~matts/coralsimplex.html`.

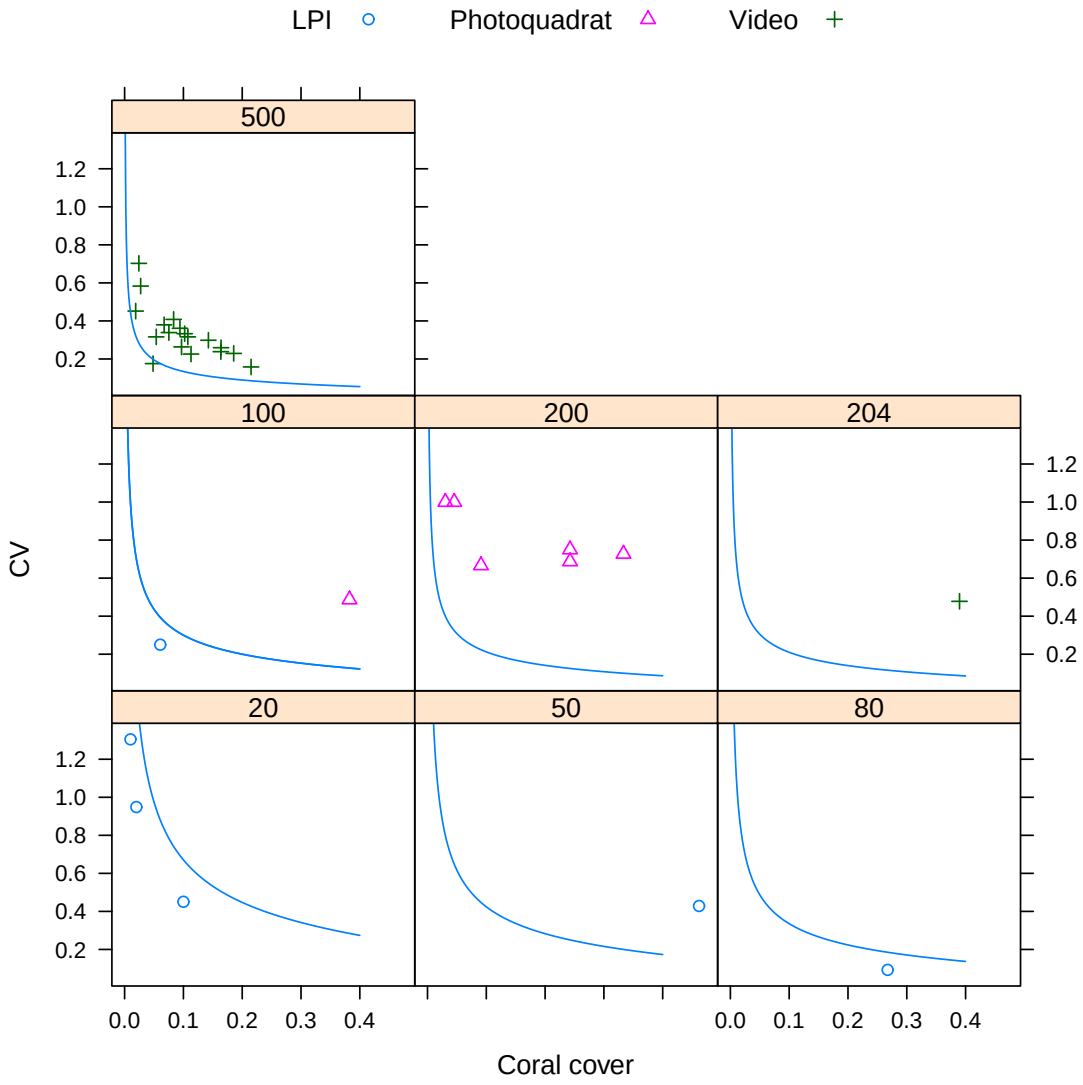

Figure S6: Relationships between coefficient of variation (CV) and mean coral cover for the data in Table S1, sorted by number of points per unit. Blue lines: CV from multinomial model (Equation S14). Blue circles: line point intercept. Pink triangles: photoquadrats. Green crosses: video transects.

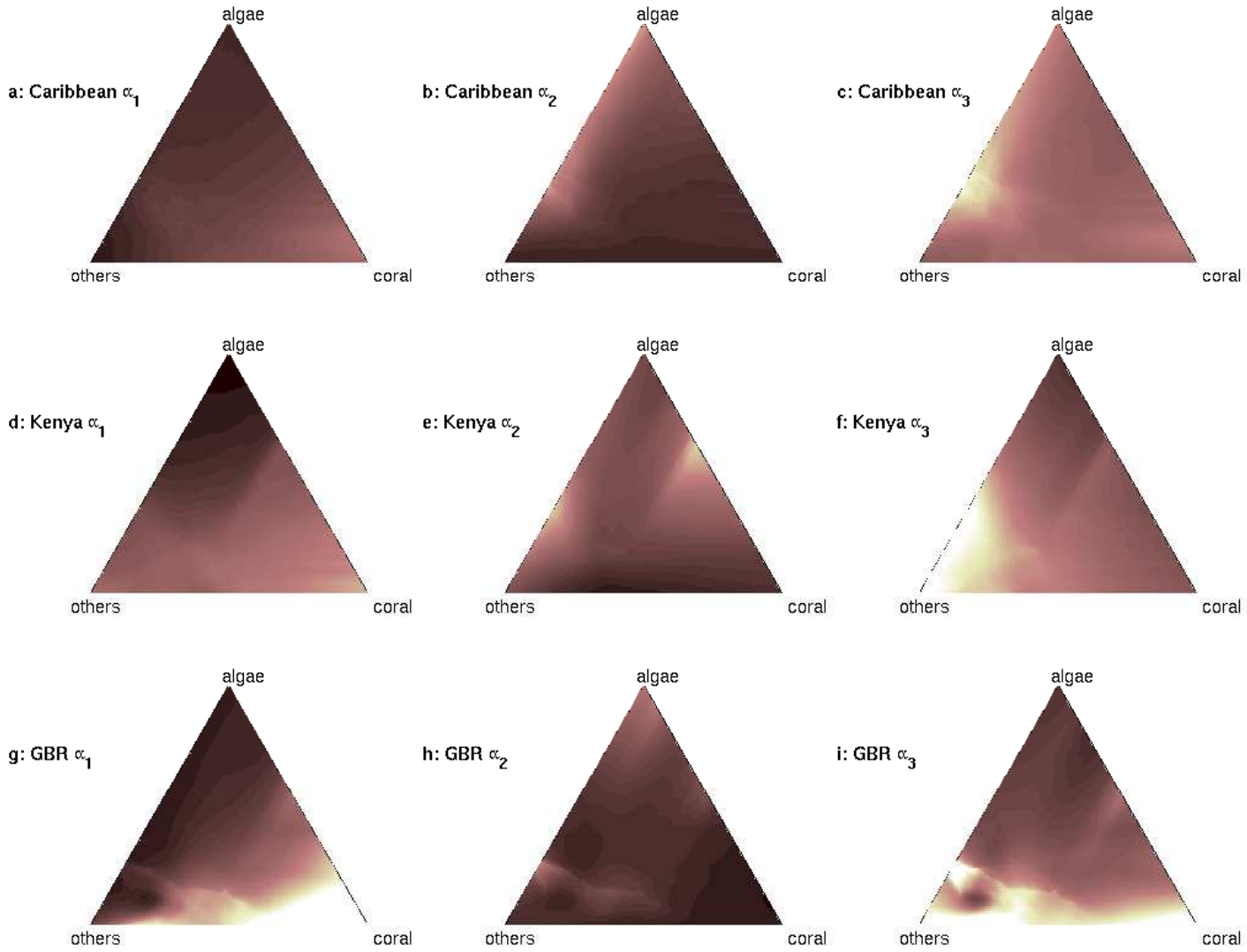

Figure S7: Local estimates of the Dirichlet $\boldsymbol{\alpha}$ parameters for each of the three regions (a-c Caribbean, d-f Kenya, g-i GBR), using proportions of neighbours (Caribbean 0.664, Kenya 0.627, GBR 0.288) selected by cross-validation as described in Section S1.6. Colour scales are truncated at 25 to show detail. In each row, α_1 is the Dirichlet parameter associated with coral, α_2 is associated with algae, and α_3 is associated with others.

S2 Supplementary results

S2.1 Estimated Dirichlet $\boldsymbol{\alpha}$

The local estimates of the $\boldsymbol{\alpha}$ parameters of the Dirichlet distribution are relatively smooth for the Caribbean (Figure S7a-c) and Kenya (Figure S7d-f). They are somewhat more bumpy for the GBR (Figure S7g-i). This is expected because the GBR is the largest data set, and will therefore allow us to see finer detail of local patterns. We think it unlikely that the fine-scale bumpiness in the GBR parameters has much impact on the pattern of trajectories. The bumpiness occurs mainly in the lower left-hand region of the simplex (Figure S7g-i). In this region, the resulting mean trajectories (main text Figure 2i) can be summarized qualitatively as "little movement in any direction". Thus, once a GBR reef has entered this region, it will tend to stay there, making this the high-density region of the equilibrium distribution.

S2.2 Simulated data

We show one typical set of simulated fates for each of the observed reef states (Figure S8). Note that because these are based on one-step predictions, even when there was a long sequence of observations, each simulated trajectory consists only of the observed state in a given year (blue circle) and a sampled fate from the model (red dot).

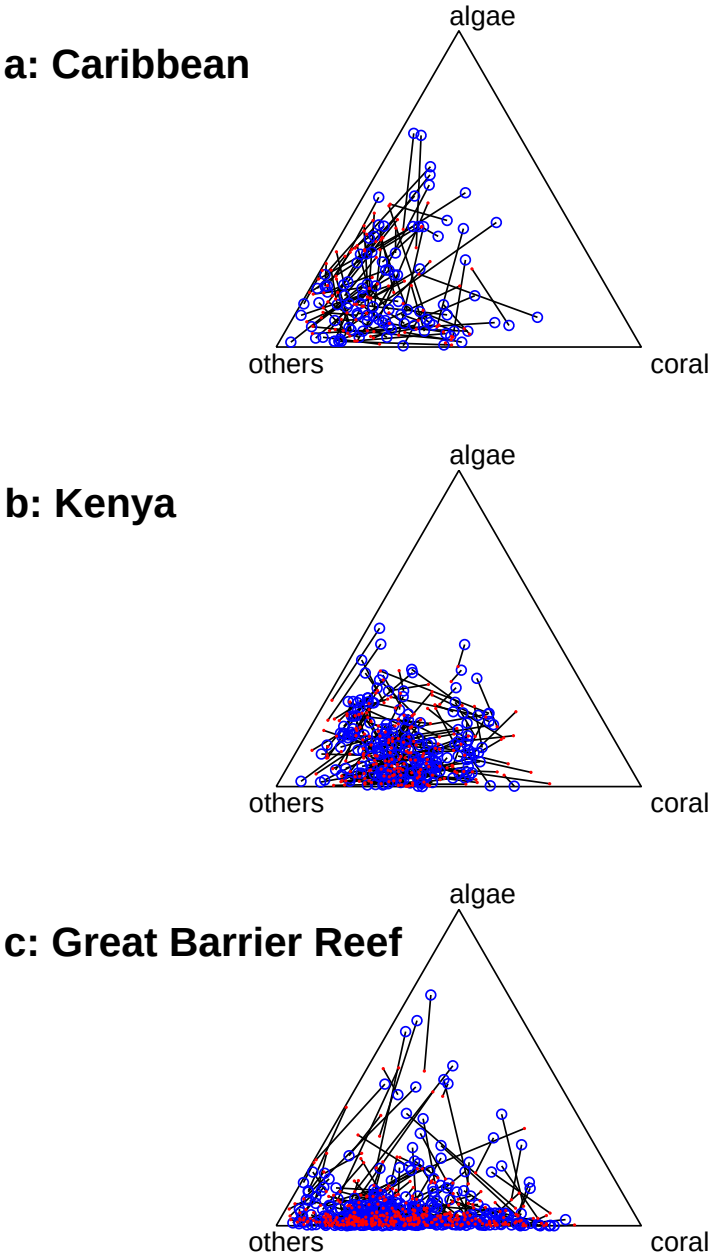

Figure S8: Simulated data under the models for (a) the Caribbean, (b) Kenya, and (c) the Great Barrier Reef. Blue circles are real observations for which there is an observation on the same reef in the next year. For each of these, the red dot connected by a black line is a single sample from the estimated transition kernel at the observation.

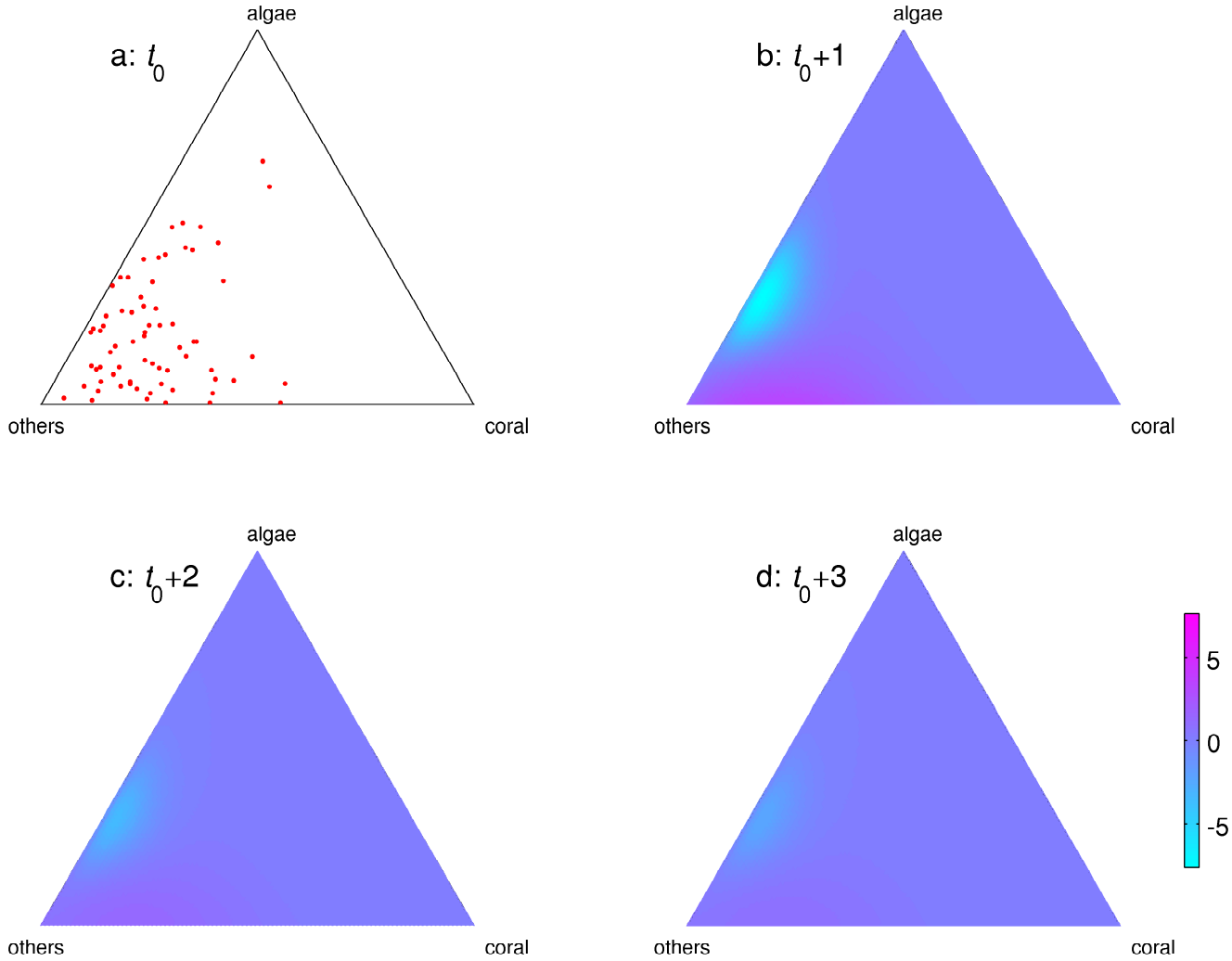

Figure S9: Predicted temporal change in the Caribbean. (a) Current distribution of reef states (red dots: most recent observation on each reef). (b), (c) and (d). Predicted differences between the estimated stationary distribution and the distribution for the points in part (a) after 1, 2, and 3 years respectively. Pink values are where the probability density after a given time is greater than in the stationary distribution, and cyan values are where the probability density is less than in the stationary distribution. The colour scales differ between this and the following two figures, in order to show detail for individual regions.

S2.3 Differences between current and equilibrium distributions

In the Caribbean (Figure S9) the absolute differences between current and equilibrium distributions are relatively small, but there are currently fewer reefs with moderate algae and low coral than predicted at equilibrium, and more with moderate coral and low algae. In Kenya (Figure S10), the absolute differences are also relatively small, but there are currently fewer reefs with moderate coral and low algae, and more with higher coral and/or algae than predicted at equilibrium. The Great Barrier Reef (Figure S11) has the largest absolute difference between current and equilibrium distributions: there are currently fewer reefs with high coral and low algae, and more with moderate coral and low algae, than at equilibrium.

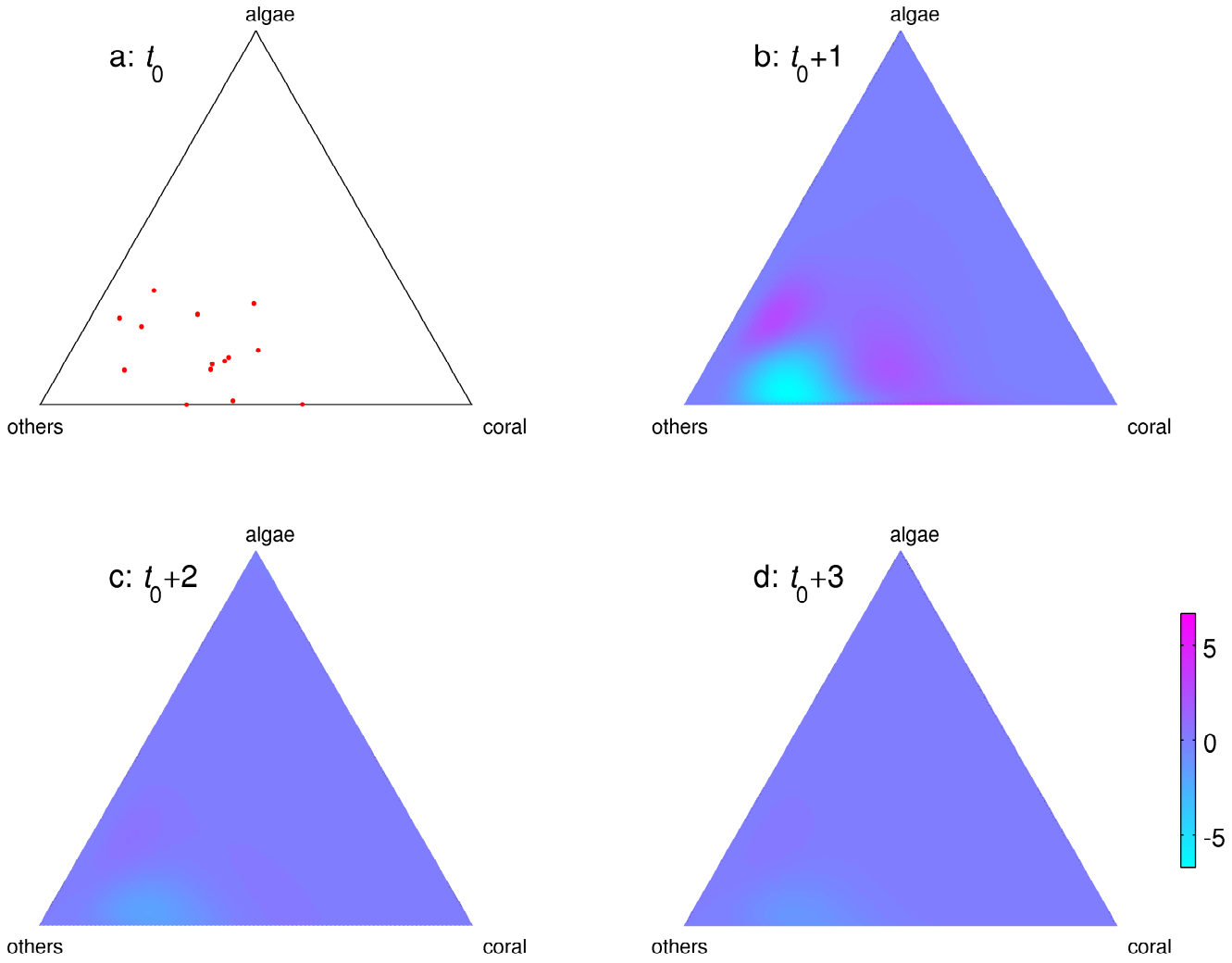

Figure S10: Predicted temporal change in Kenya. See Figure S9 for details.

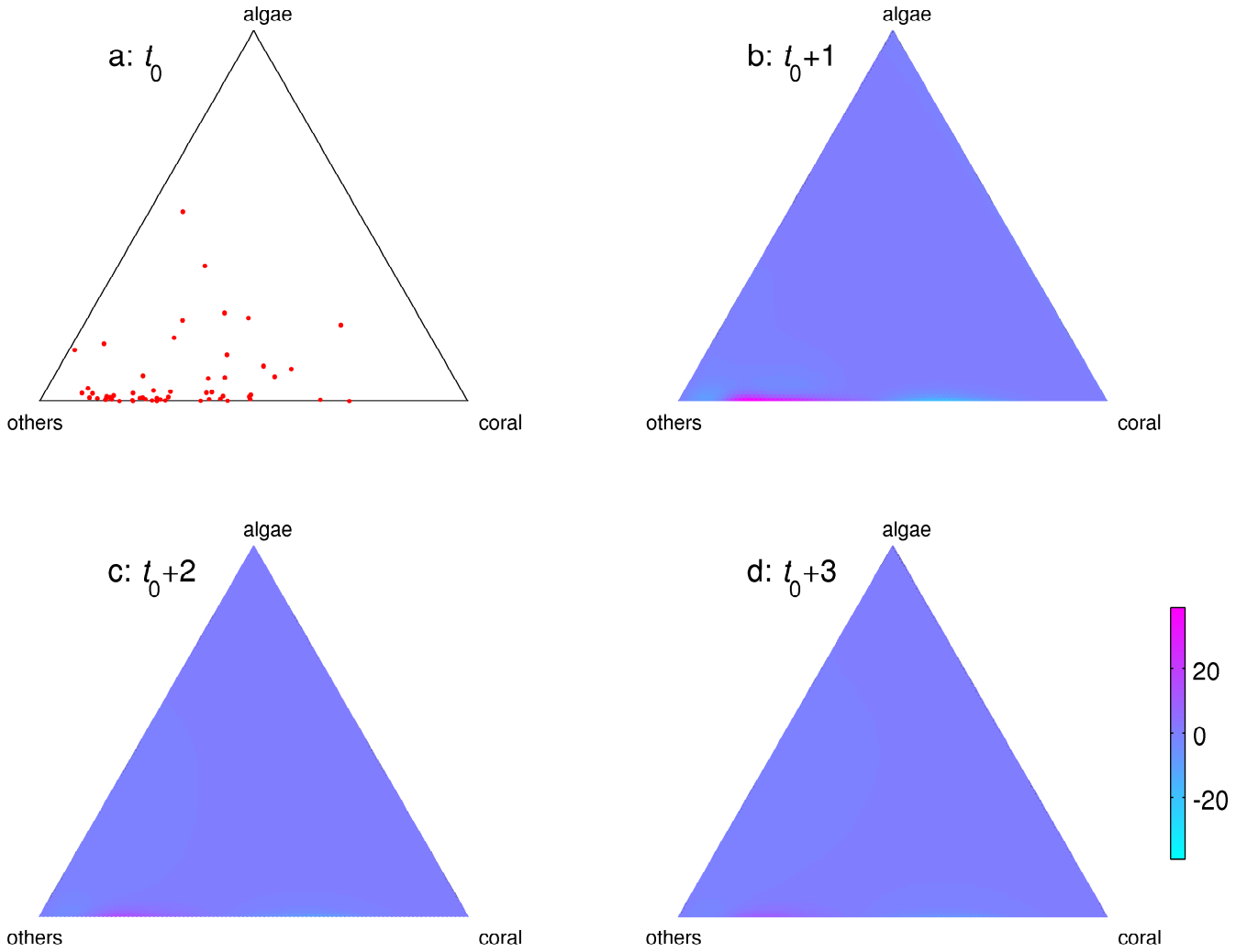

Figure S11: Predicted temporal change in the Great Barrier Reef. See Figure S9 for details.

Table S2: Sample means and standard deviations of standardized univariate residuals. Each entry is mean, standard deviation of the standardized difference between the observed and predicted proportion of each category (Section S1.10). Thus, although both observed and predicted proportions are constrained to be between 0 and 1, the residuals may be either negative (if there is less of a category than predicted) or positive (if there is more).

Dataset	Coral	Algae	Others
Caribbean	-0.11, 0.67	0.07, 1.22	0.001, 1.12
Kenya	0.03, 0.88	0.09, 1.03	-0.10, 0.88
GBR	-0.07, 0.80	0.13, 1.37	-0.02, 0.93

S2.4 Residuals

Plots of residuals against time suggested that there were some unusual years. In the Caribbean (Figure S12), 2005 had unusually high algae and low others. This could be either an unusual event, or an unusual subset of reefs sampled. In Kenya (Figure S13), 1999 had unusually low coral, high algae, and high others (because the sampling date changed from December to January in 1998/99, this corresponds to events occurring during 1998). In the GBR (Figure S14), there were six reefs with outlying large composite residuals in 1998, with high algae and/or low others. Both these patterns may be associated with the 1998 El Niño, as discussed in the Results section of the main text.

If the model accurately described the data, we would expect the standardized residuals for each component to have mean approximately 0 and standard deviation approximately 1. In fact (Table S2), the mean residuals in the Caribbean and GBR tend to be negative for coral and positive for algae. In Kenya, the mean residuals tend to be negative for others and positive for algae. These biases are small relative to the amount of variability. However, in all regions, algal residuals are more variable than expected, perhaps because of rapid but short-lived algal increases. The standardized residuals for others are also more variable than expected in the Caribbean, and less variable than expected in the GBR and Kenya.

For the Caribbean (Figure S15) and Kenya (Figure S16), there are no strong patterns in the residuals for fate with respect to current reef state (the year of the observation whose fate we want to predict). The only strong relationship between current reef state and residual for fate is that in the Great Barrier Reef, residuals for all three components tend to be more variable when algal cover is low than high (Figure S17d, e, and f). A partial explanation for this is that we have many observations with low algal cover, spanning a wide range of coral cover. In contrast, there are few observations with high algal cover, all having low coral cover (Main text, Figure 2c). It is possible that the reefs with high algal cover are more homogeneous in physical conditions than those with low algal cover.

The Caribbean residuals tend to be more variable when algal cover in the year before the observation is low than when it is high (Figure S18d, e, and f). For Kenya, there are no strong patterns in the residuals with respect to reef state the year before the observation whose fate we want to predict (Figure S19). The pattern for the Great Barrier Reef is similar to that for the Caribbean, but stronger (Figure S20d, e and f). The pattern in the Caribbean and Great Barrier Reef could indicate a violation of the Markovian assumption, or it could simply mean that reef states tend to be similar in adjacent years (because we do not have an additive time series model, taking residuals does not condition on current state).

There was a tendency for the Caribbean reefs with the highest composite residuals to have more algae and less others than predicted (Figure S21). Thus, when our predictions in the Caribbean are least accurate, they show a consistent bias. However, this tendency is driven by six reefs, for each of which we have only one state and one fate. There is relatively little evidence of patterns in residuals by reef in Kenya (Figure S22) or the Great Barrier Reef (Figure S23). We discuss these patterns in detail in the next section, and show that omitting possibly unusual reefs does not change our conclusions.

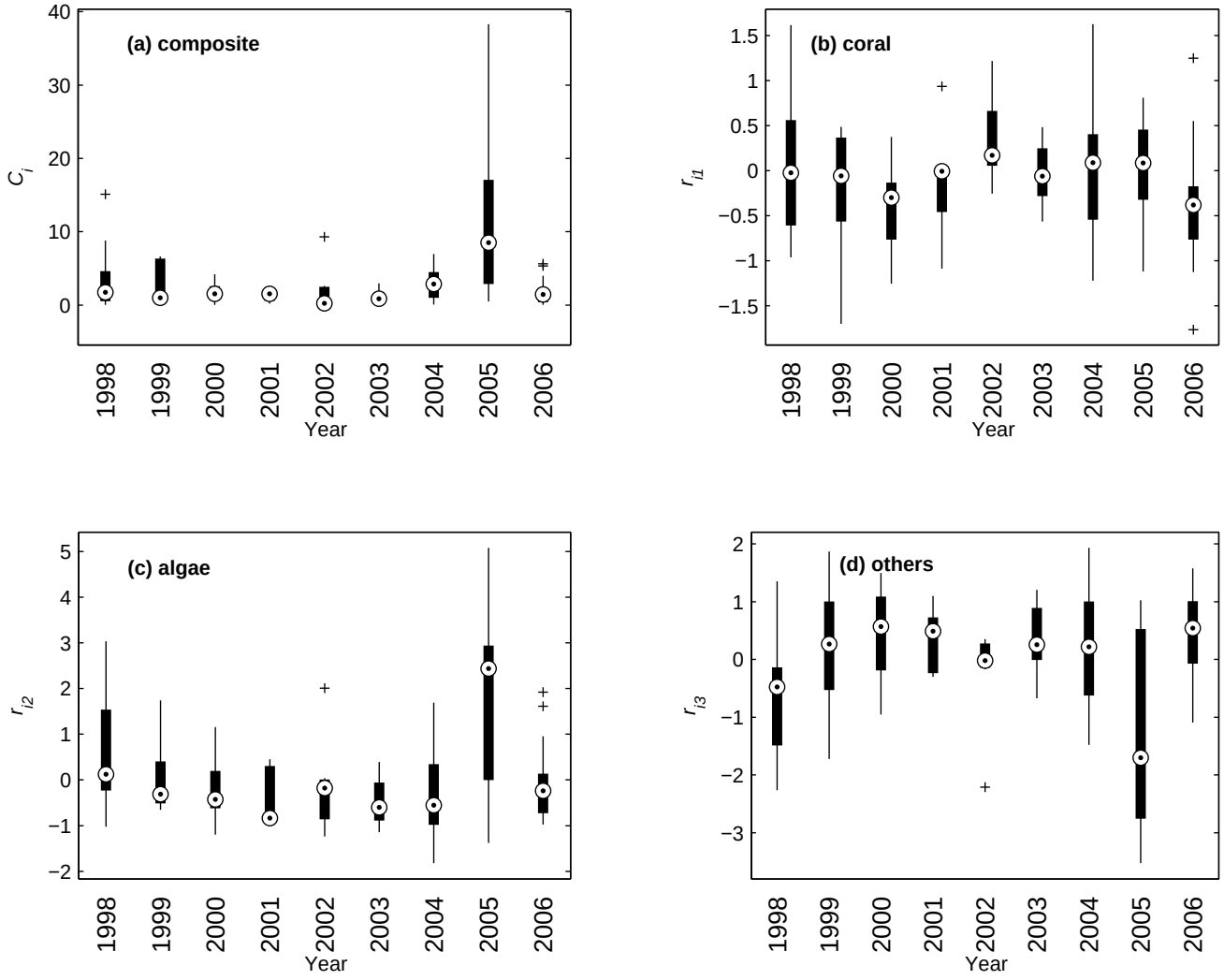

Figure S12: Boxplots of composite residuals C_i (a) and univariate standardized residuals r_{ij} (b: coral, $j = 1$, c: algae, $j = 2$, d: others, $j = 3$) for each observation i, grouped by time for the Caribbean. Year is the date of the observation we are predicting. The vertical scales are different for each panel.

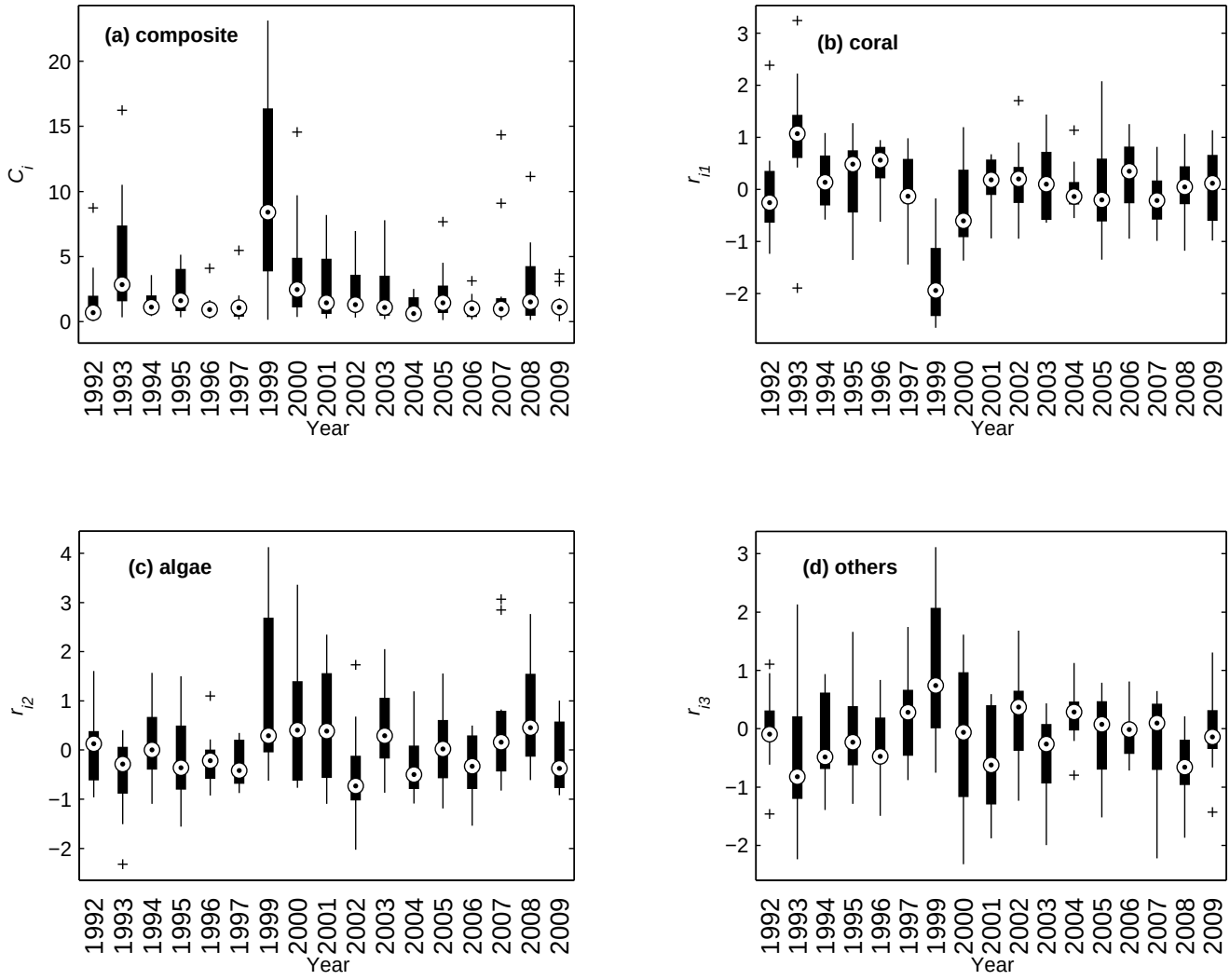

Figure S13: Boxplots of composite residuals C_i (a) and univariate standardized residuals r_{ij} (b: coral, $j = 1$, c: algae, $j = 2$, d: others, $j = 3$) for each observation i, grouped by time for Kenya. Year is the date of the observation we are predicting. 1998 is omitted because even though observations were made at roughly 12-month intervals, they were made in December up until 1997, then in January from 1999 onwards. The vertical scales are different for each panel.

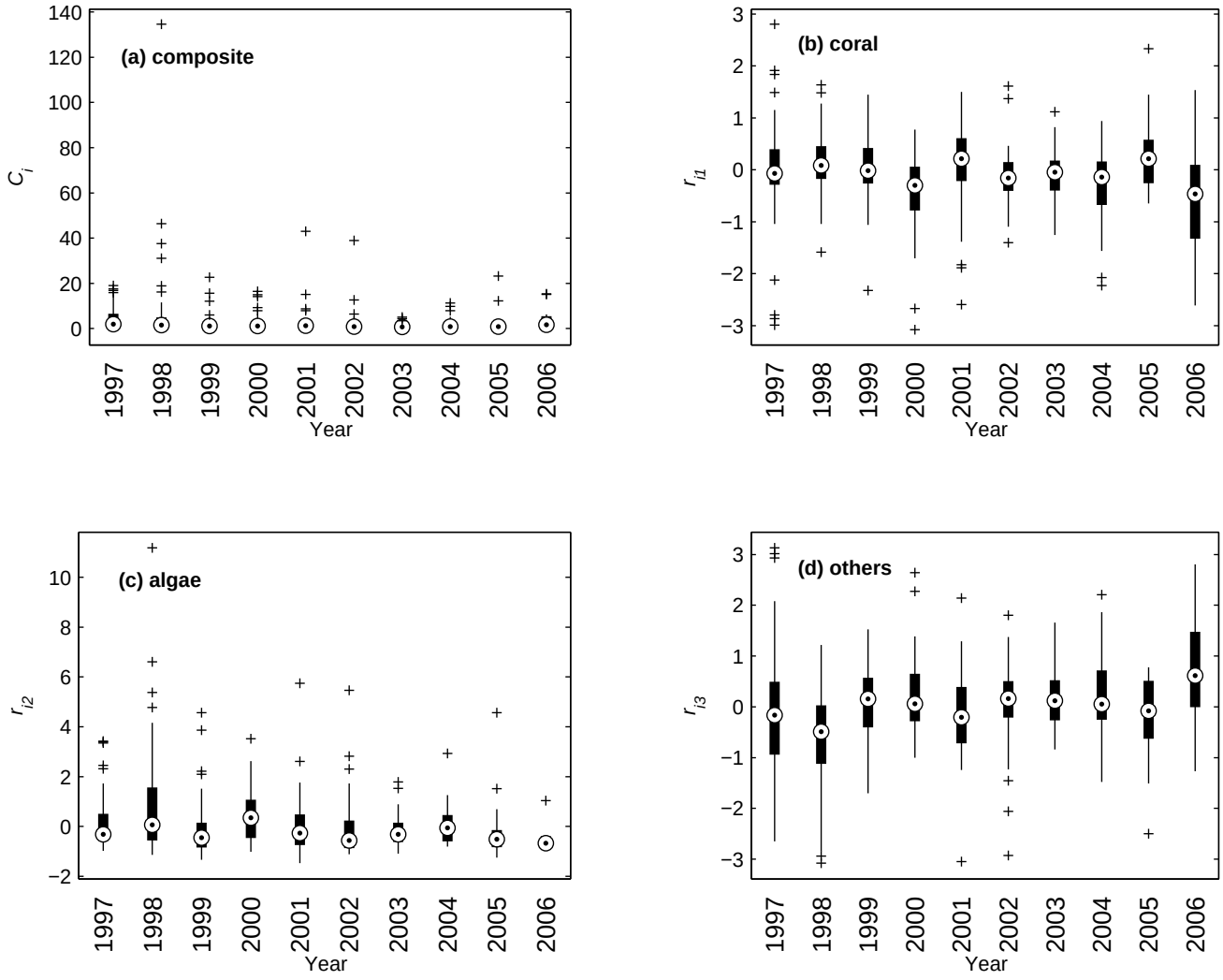

Figure S14: Boxplots of composite residuals C_i (a) and univariate standardized residuals r_{ij} (b: coral, $j = 1$, c: algae, $j = 2$, d: others, $j = 3$) for each observation i, grouped by time for the GBR. Year is the date of the observation we are predicting. The vertical scales are different for each panel.

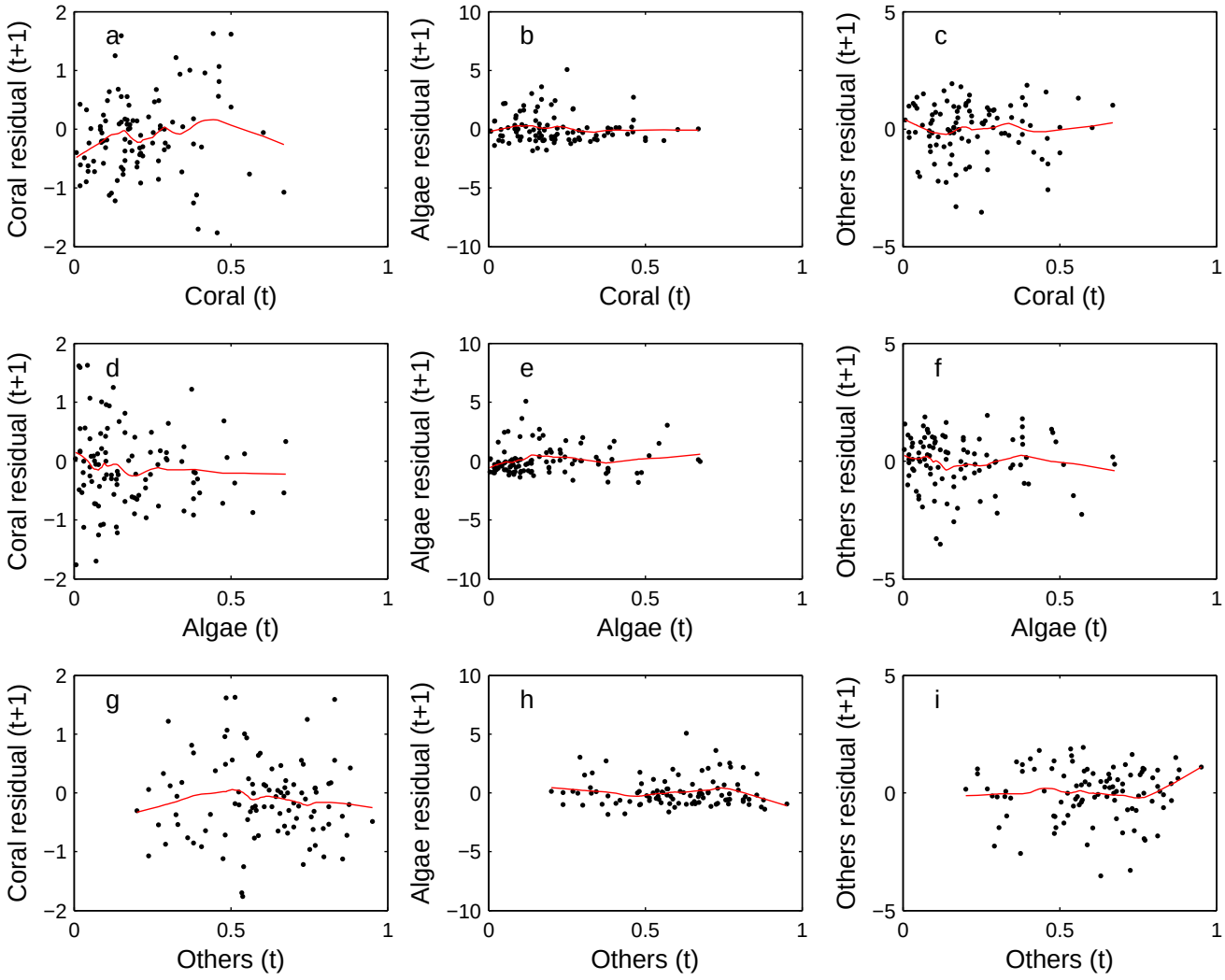

Figure S15: Standardized univariate residuals for Caribbean data against composition in the observation $\mathbf{x}$. Points are values of standardized residuals for coral (a, d, g), algae (b, e, h) and others (c, f, i) against the cover of coral (a, b, c), algae (d, e, f), and others (g, h, i) in the observation $\mathbf{x}$. The red lines are LOWESS smoothers with span 0.4.

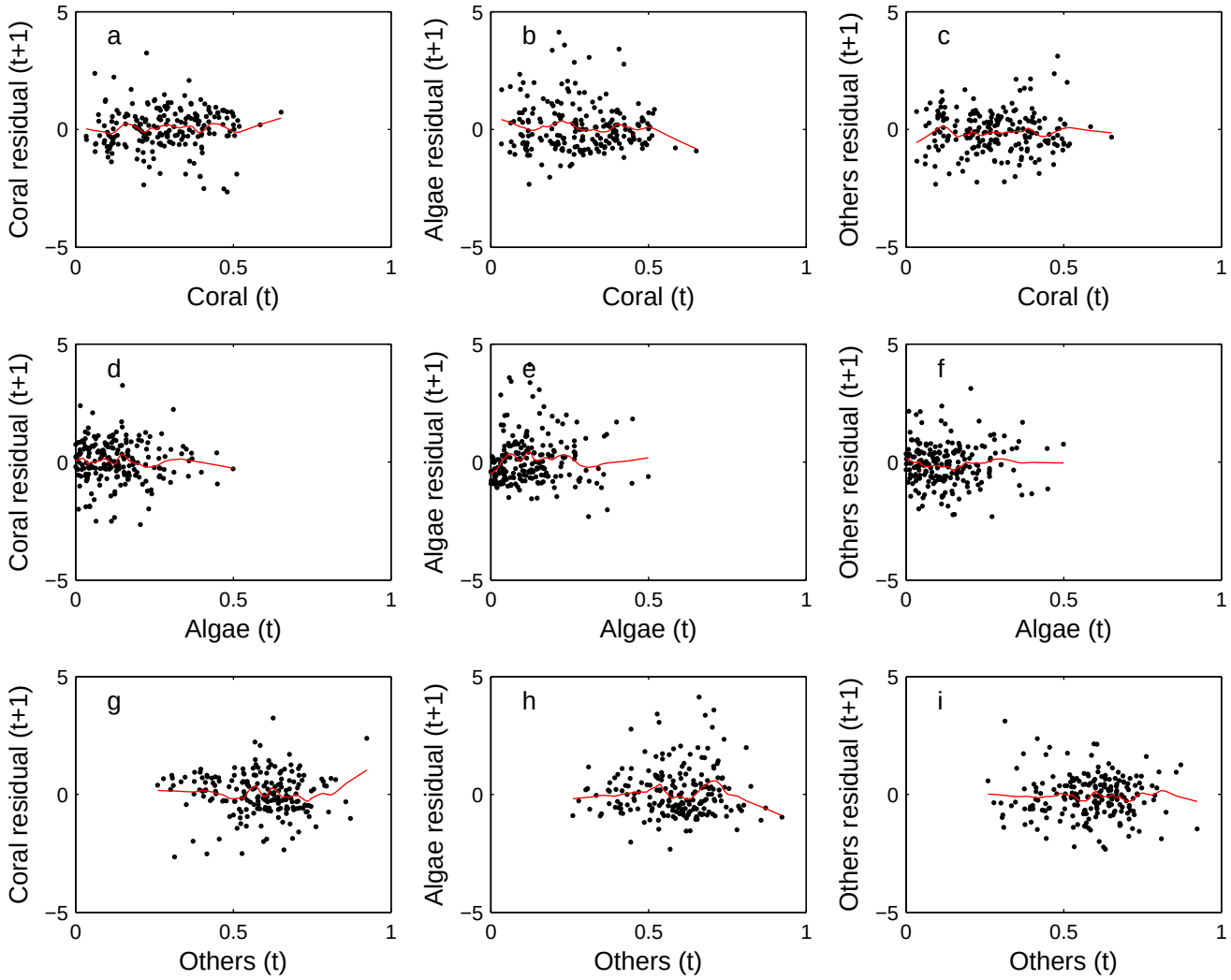

Figure S16: Standardized univariate residuals for Kenyan data against composition in the observation $\mathbf{x}$. Points are values of standardized residuals for coral (a, d, g), algae (b, e, h) and others (c, f, i) against the cover of coral (a, b, c), algae (d, e, f), and others (g, h, i) in the observation $\mathbf{x}$. The red lines are LOWESS smoothers with span 0.2.

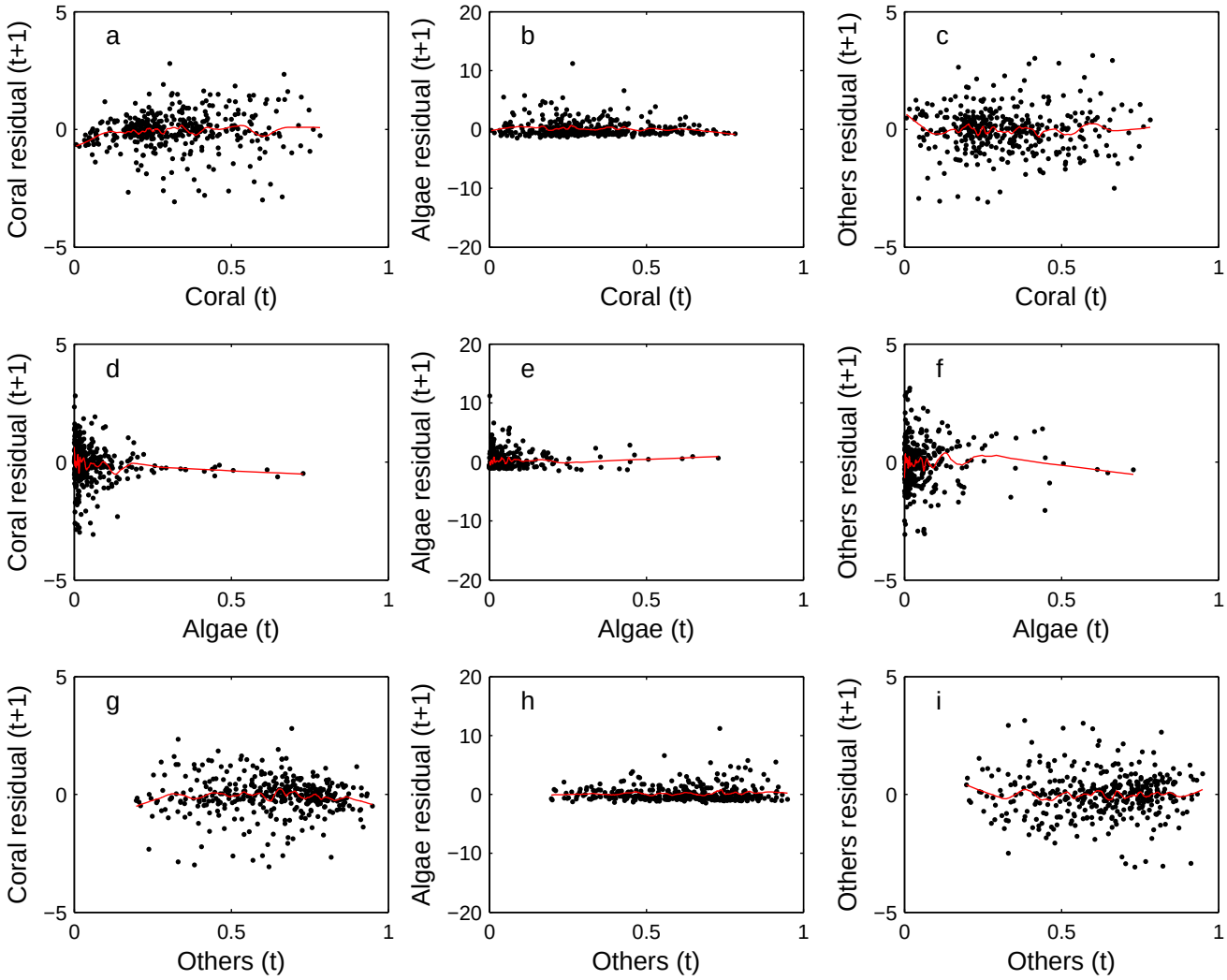

Figure S17: Standardized univariate residuals for Great Barrier Reef data against composition in the observation $\mathbf{x}$. Points are values of standardized residuals for coral (a, d, g), algae (b, e, h) and others (c, f, i) against the cover of coral (a, b, c), algae (d, e, f), and others (g, h, i) in the observation $\mathbf{x}$. The red lines are LOWESS smoothers with span 0.1.

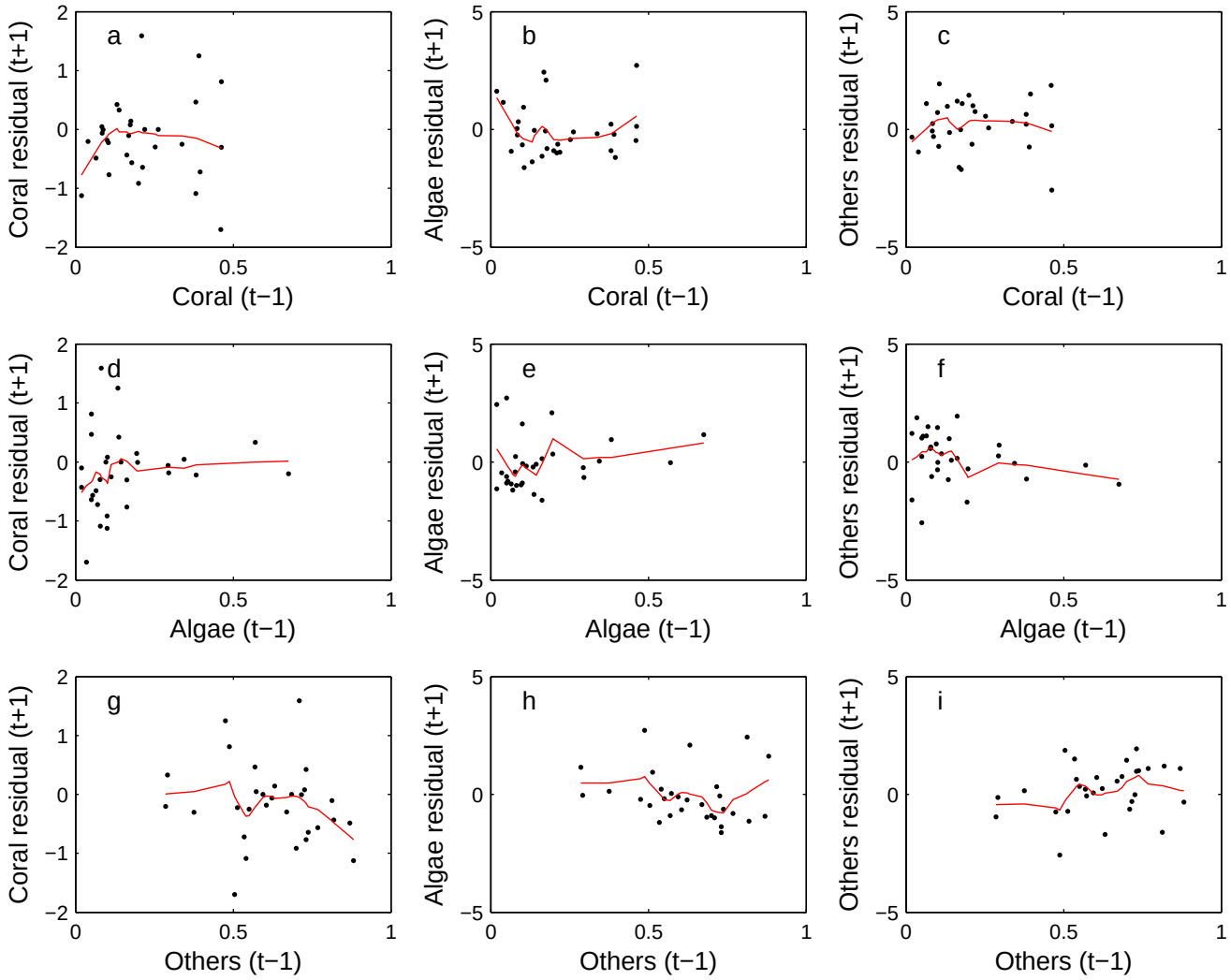

Figure S18: Standardized univariate residuals for Caribbean data against composition the year before the observation $\mathbf{x}$. Points are values of standardized residuals for coral (a, d, g), algae (b, e, h) and others (c, f, i) against the cover of coral (a, b, c), algae (d, e, f), and others (g, h, i) the year before the observation $\mathbf{x}$. The red lines are LOWESS smoothers with span 0.4.

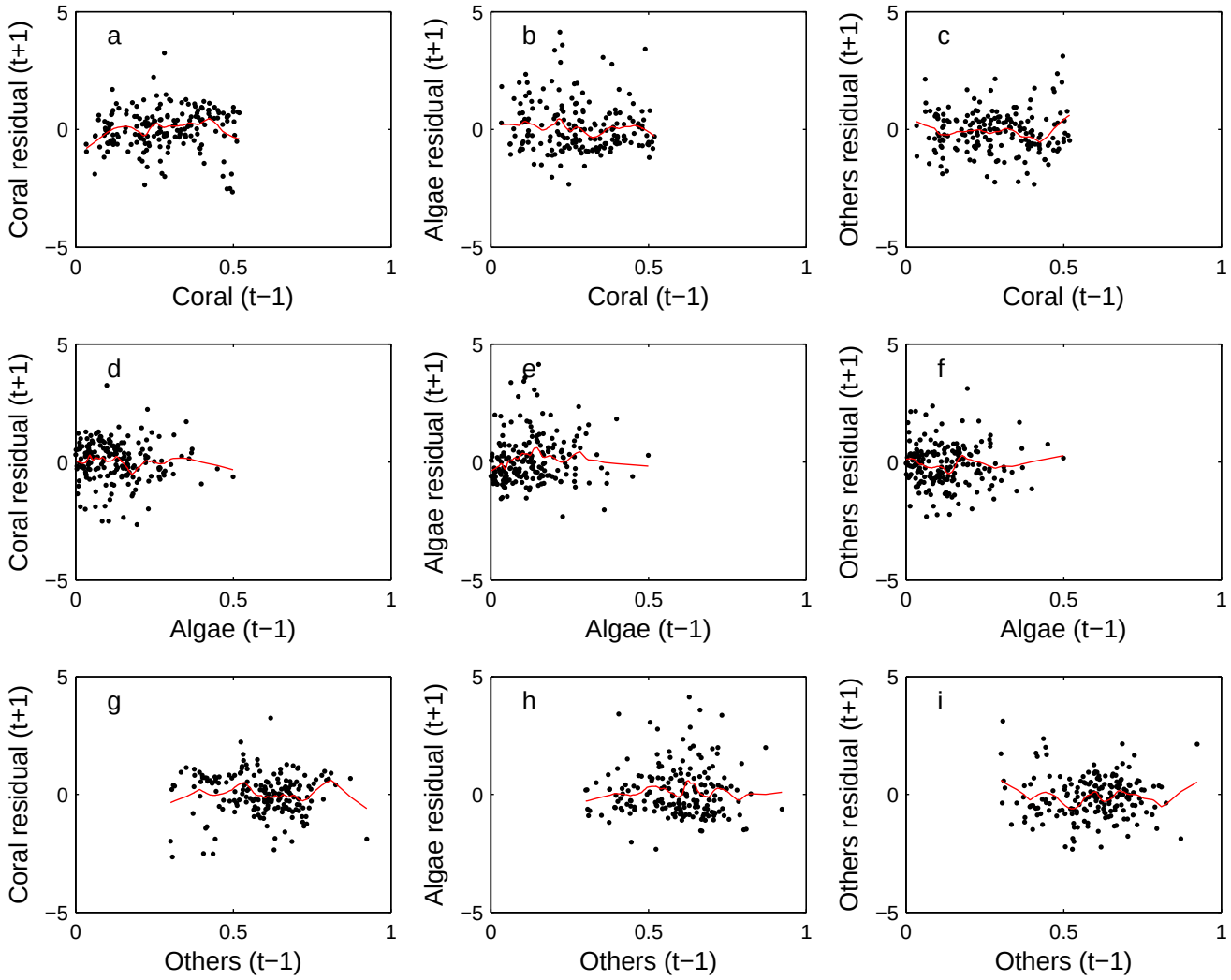

Figure S19: Standardized univariate residuals for Kenyan data against composition the year before the observation $\mathbf{x}$. Points are values of standardized residuals for coral (a, d, g), algae (b, e, h) and others (c, f, i) against the cover of coral (a, b, c), algae (d, e, f), and others (g, h, i) the year before the observation $\mathbf{x}$. The red lines are LOWESS smoothers with span 0.2.

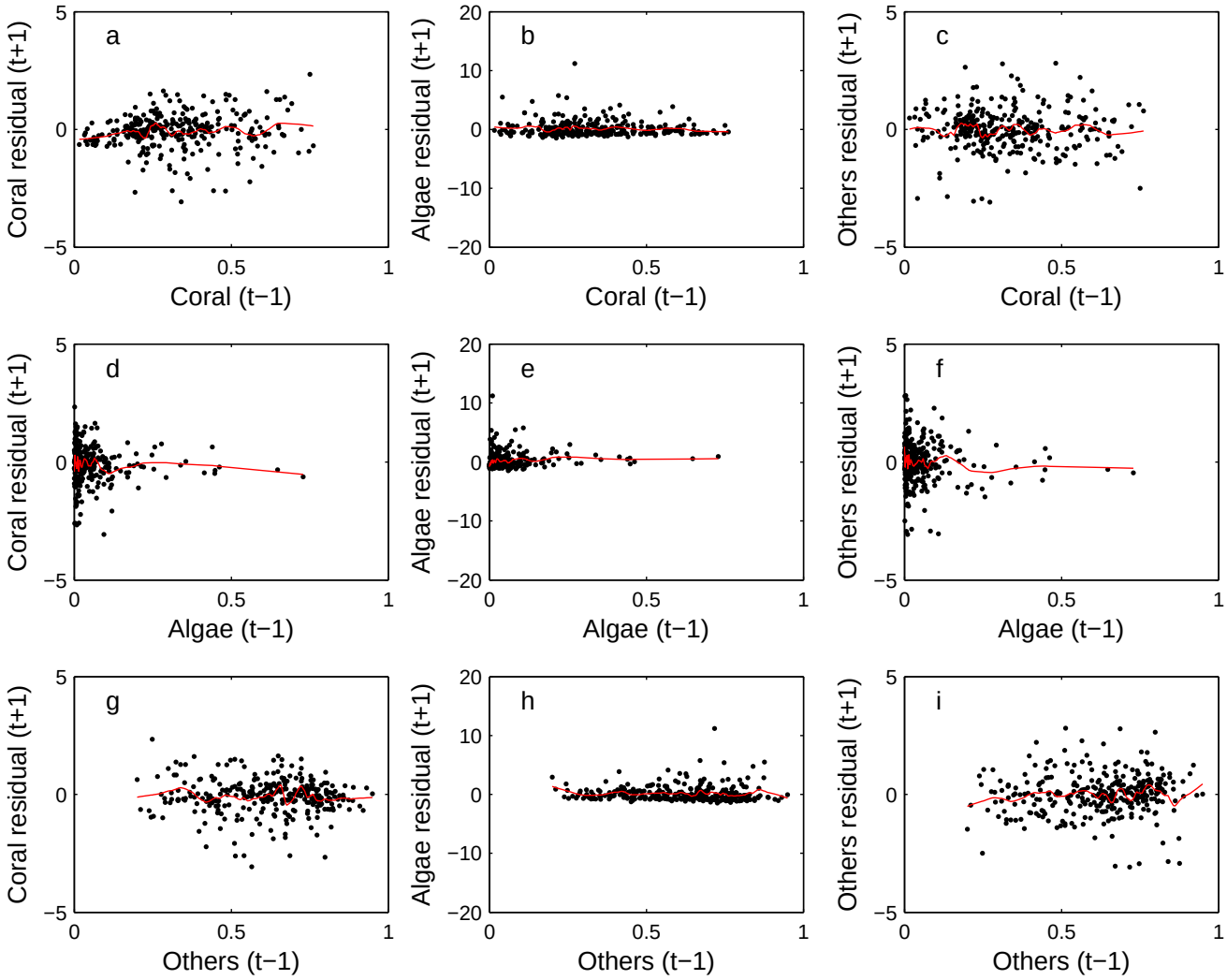

Figure S20: Standardized univariate residuals for Great Barrier Reef data against composition the year before the observation $\mathbf{x}$. Points are values of standardized residuals for coral (a, d, g), algae (b, e, h) and others (c, f, i) against the cover of coral (a, b, c), algae (d, e, f), and others (g, h, i) the year before the observation $\mathbf{x}$. The red lines are LOWESS smoothers with span 0.1.

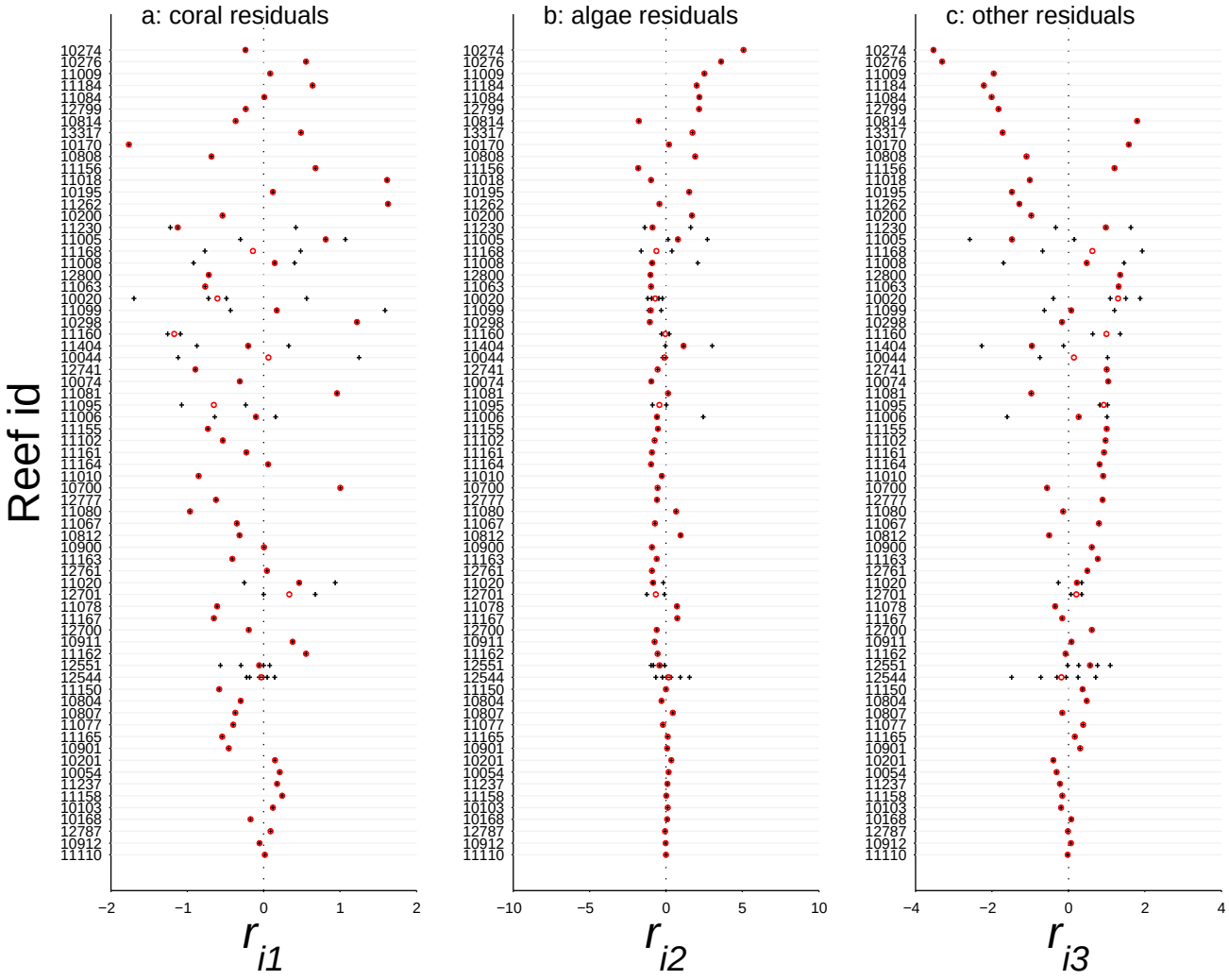

Figure S21: Standardized univariate residuals for Caribbean data by reef. For each component (a: coral, b: algae, c: others), the residuals for each reef (row) are shown as black crosses, with a red circle at the median. Reefs are identified by numerical codes and sorted from top to bottom in descending order of median composite residual, where the composite residual is the sum of the squared standardized univariate residuals.

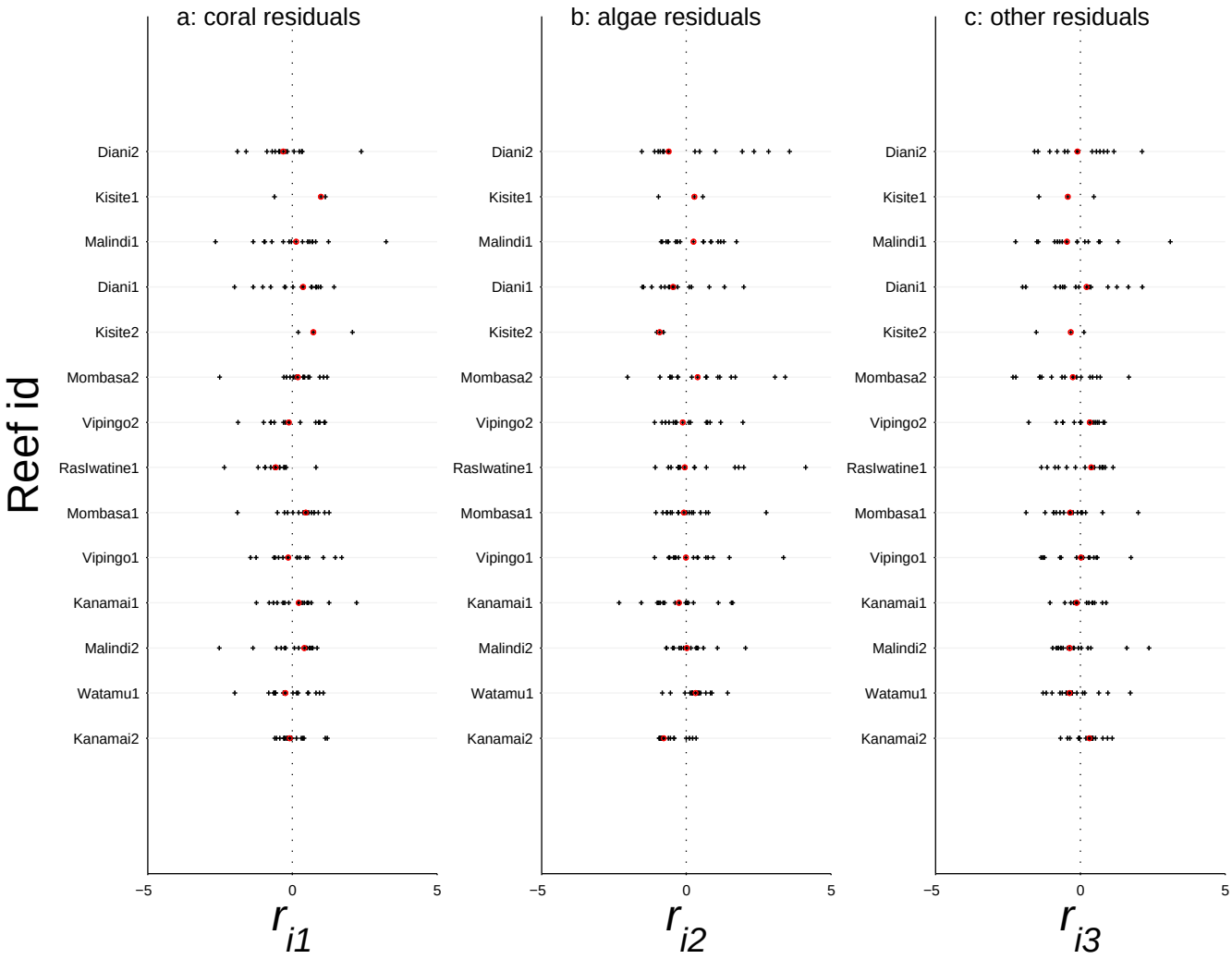

Figure S22: Standardized univariate residuals for Kenya data by reef. See Figure S21 for explanation. Reefs in this region are identified by name.

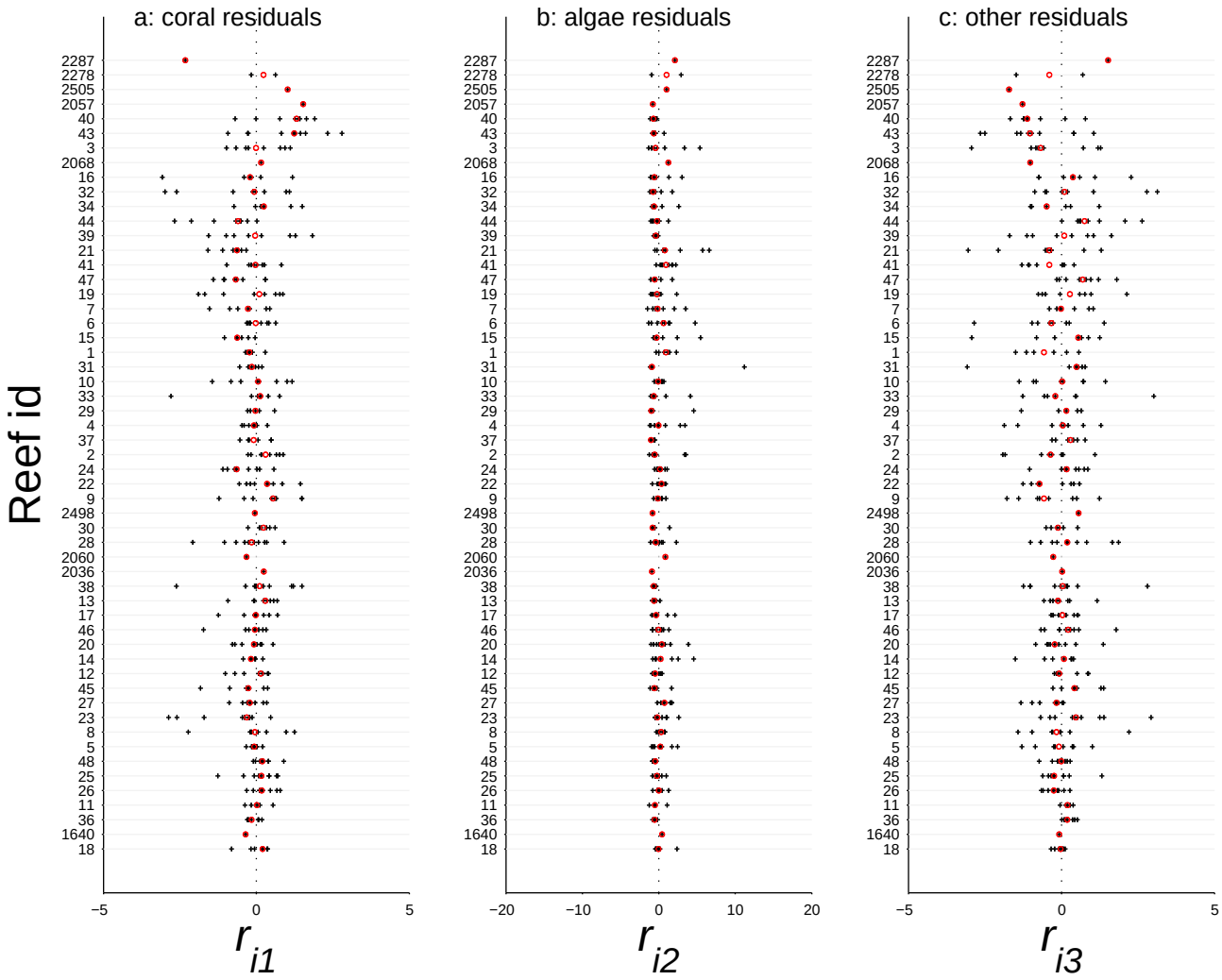

Figure S23: Standardized univariate residuals for Great Barrier Reef data by reef. See Figure S21 for explanation.

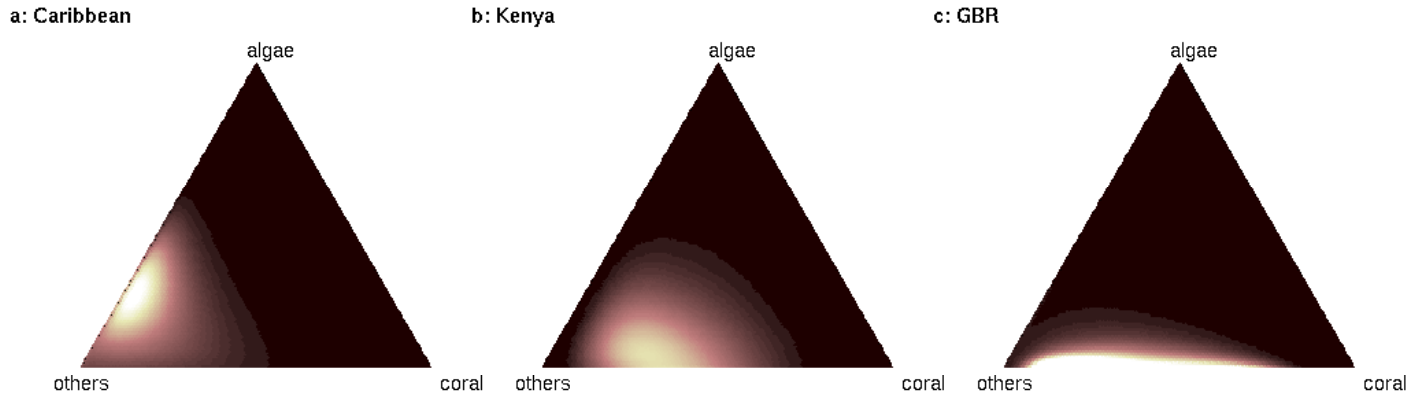

Figure S24: Estimated equilibrium distributions for (a) Caribbean (b) Kenya, and (c) GBR, omitting reefs with potentially unusual patterns of residuals. Lighter colours are higher probability densities, with a colour scale truncated at 25, as in the main text

S2.5 Re-analysis without unusual reefs

A few reefs have patterns of residuals that might indicate systematic differences from the overall pattern. In the Caribbean (Figure S21), there are relatively few reefs with more than two observations, so it is not easy to detect reefs with consistently unusual residuals. Nevertheless, there is one reef (12551 US Virgin Islands) with four out of five positive residuals for the others component. In addition, the six reefs with the largest median composite residuals (10274 Havana Cuba, 10276 Havana Cuba, 11009 Dominican Republic, 11184 Barbados, 11084 US Virgin Islands, 12799 US Virgin Islands) all show the same pattern of positive algae residuals and negative other residuals, although there is only one residual for each.

In Kenya (Figure S22), RasIwatine1 has all but one out of 15 coral residuals negative (and has the highest erect algal cover of all the reefs in the Kenyan dataset). In addition, although Kisite2 has only three residuals, all are positive for coral and negative for algae. This site may be unusual in having a number of very large *Galaxea astreata* colonies, which can reach up to 3m in length (T. McClanahan, personal observation). Where a transect falls in relation to these colonies could have a large influence on estimated coral cover. This high spatial variability could possibly account for the unusual residuals at Kisite2.

In the GBR (Figure S23), reef 43 has seven out of 10 residuals positive for coral, nine out of 10 negative for algae, and seven out of 10 negative for others. Reef 44 has seven out of eight residuals negative for coral, six out of eight negative for algae, and all eight positive for others. Reef 21 has all eight residuals negative for coral, six out of eight positive for algae, and six out of eight negative for others. Reef 15 has all seven residuals negative for coral. Reef 23 has nine out of 10 residuals negative for coral, and seven out of 10 positive for others.

We therefore refitted the models for each region with all these reefs omitted. The estimated equilibrium distributions (Figure S24) are little different from those based on all the data (main paper, Figure 2j-l).

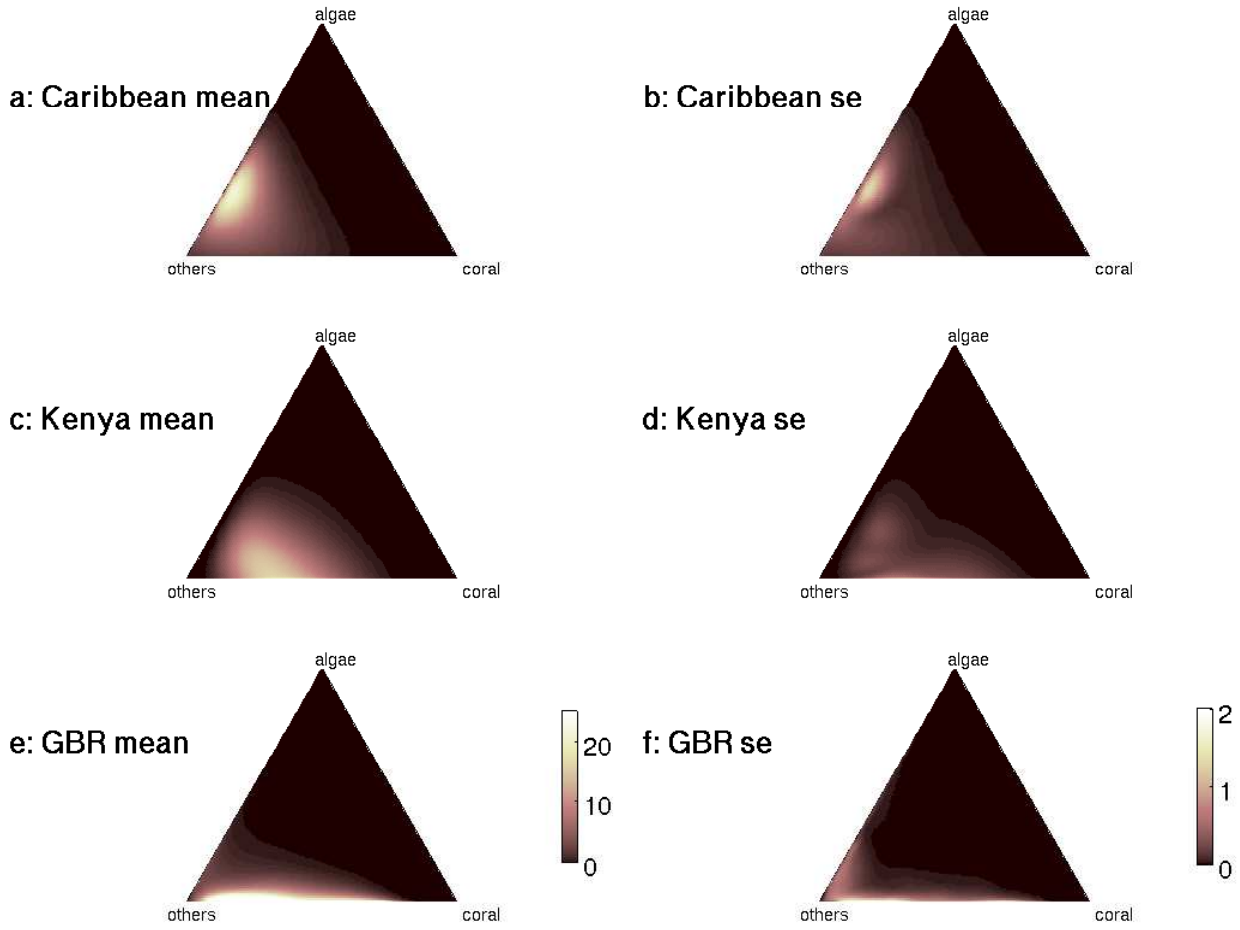

Figure S25: Jacknife means (a, c, e) and standard errors (b, d, e) of the equilibrium distributions for the Caribbean (a, b), Kenya (c, d) and GBR (e, f). Lighter colours are higher values. In a, c, and e, the colour scale is truncated at 25, as in the main text. In b, d, and f, the maximum of the colour scale is at 2 (just above the largest jacknife standard error at any grid point).

S2.6 Estimates of uncertainty

The jackknife mean estimates of the equilibrium distributions (Figure S25a, c, e) are very similar to the estimates from the full data set (main text, Figure 2j, k, l). The jackknife standard errors are relatively small, and roughly proportional to the jackknife means (Figure S25b, d, f: note the maximum of the colour scale in these panels is much less than in panels a, c, e). In the Caribbean and Kenya, all the jackknife replicates had a similar appearance to the overall mean. In the GBR, one out of 374 jackknife replicates was bimodal, with a secondary mode on the others-algae edge, close to the others corner. Overall, the uncertainty in our estimated equilibrium distributions is relatively small, and there is little support for equilibrium distributions other than the one presented in the main text.

Table S3: Effects of multinomial measurement error on the shapes of estimated stationary distributions. Sampling effort m is the number of points in the multinomial sample.

m	Unimodal	Bimodal	Trimodal	Other
20	0	0	93	7
160	1	6	41	52
1000	4	56	0	40
5000	7	79	11	2

S2.7 Estimation from a model with alternative stable states

Out of 100 replicate simulations without measurement error based on the Caribbean data, with parameters such that the true stationary distribution was bimodal, 86 of the estimated stationary distributions were bimodal and 14 unimodal. The mean and standard error of these stationary distributions are shown in Figure S26. In the unimodal cases, the single mode roughly corresponded to one of the two true modes.

S2.8 Consequences of measurement error

The effects of multinomial measurement error on the estimated stationary distribution, in a case where the true stationary distribution was bimodal, depended strongly on sampling effort. With a very low effort of 20 points, the mean estimated stationary distribution was strongly trimodal (Figure S27), with high density in the corners of the simplex. With an effort of 160 points (used by Reef Check), the mean estimated stationary distribution still had high density in the corners, but was more diffuse (Figure S28). The substantial decrease in the mean perturbation norm when Reef Check data were omitted from the Caribbean (main text, Table 1) supports the idea that most of our data had higher sampling effort than Reef Check. With higher sampling effort of 1000 points, the mean estimated stationary distribution had a strongly bimodal appearance, although the others=1 corner also had a small high density region (Figure S29). With the highest simulated effort of 5000 points, recommended by Aronson et al. (1994), the mean estimated stationary distribution was bimodal (Figure S30). As effort increased, the proportion of replicates in which the estimated stationary distribution was bimodal also increased (Table S3). However, low effort tended to result in trimodal or other (typically high density in the corners and at an internal mode) patterns, rather than unimodal distributions.

Estimated stationary distributions tended to be trimodal at low sampling effort because of the relationship between measurement error and reef composition (Figure S31). For states in which the three parts are fairly equally represented, measurement error spreads the estimated states out much more than for states lying close to an edge or in a corner of the simplex. As a result, measurement error tends to concentrate estimated fates in the corners of the simplex more than in the centre, and in the long run, this leads to stationary distributions with high density in the corners.

Measurement error clearly has the potential to affect estimated stationary distributions at plausible sampling effort. Nevertheless, our estimated stationary distributions (main text, Figure 2j-l) do not show the trimodal pattern characteristic of a strong influence of measurement error. In addition, omitting the Reef Check data, likely to have the lowest sampling effort, from the Caribbean and Great Barrier Reef regions, made very little difference to the qualitative shapes of the estimated stationary distributions (Figure S32). We therefore conclude that measurement error is unlikely to have been responsible for us failing to detect bimodal stationary distributions in the real data.

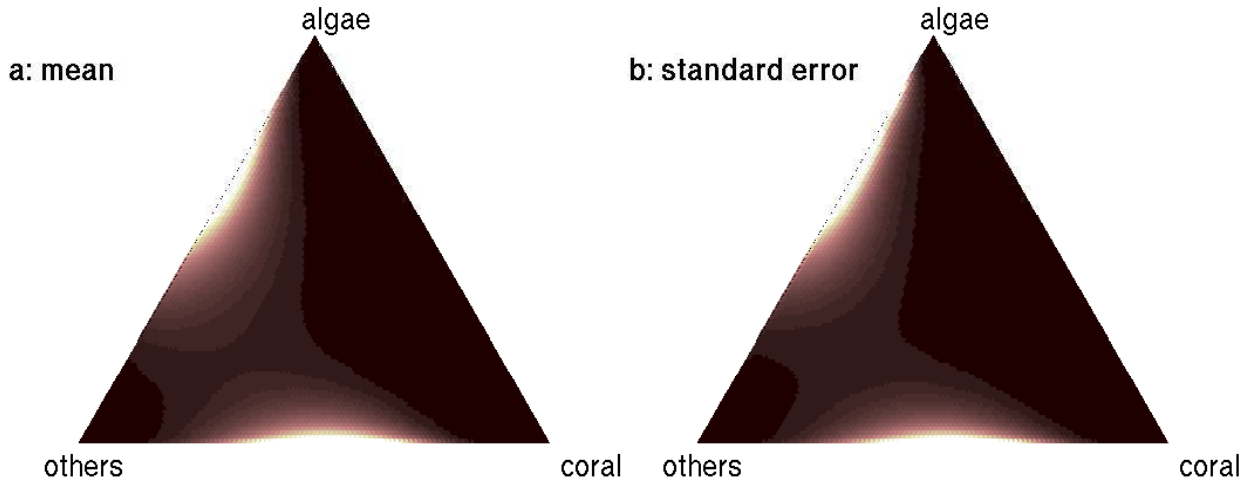

Figure S26: Mean (a) and standard error (b) of the stationary distributions of 100 replicates simulated under the model used to produce Figures S4 and S5, without measurement error. Lighter colours are higher probability densities. The maximum of the colour scale is set to 25, as for the real data, although the highest mean density is 72 and the highest standard error of density is 75.
.

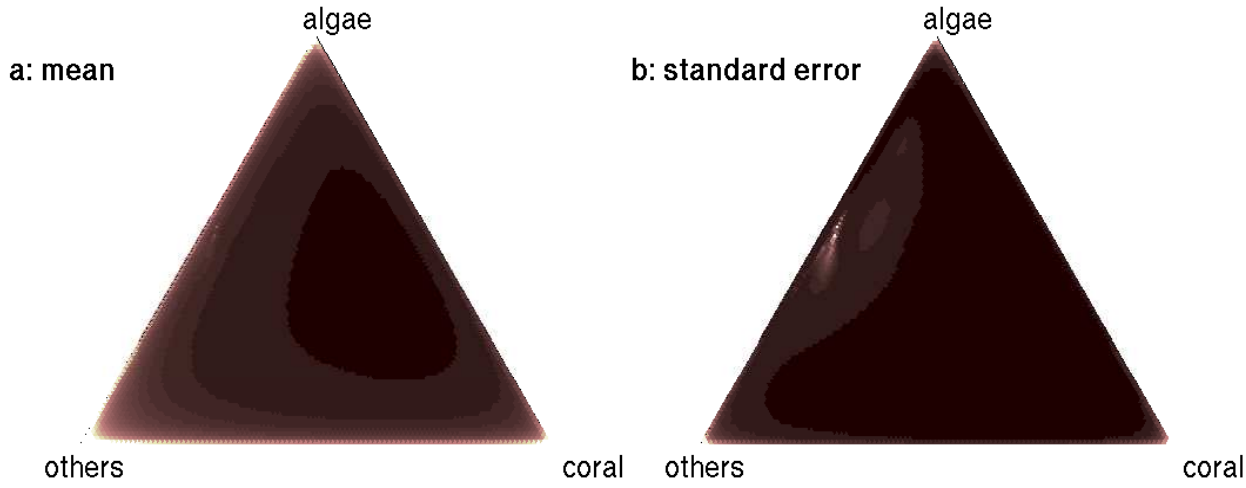

Figure S27: Mean (a) and standard error (b) of the stationary distributions of 100 replicates simulated under the model used to produce Figures S4 and S5, with multinomial measurement error, $m = 20$ points. Lighter colours are higher probability densities. The maximum of the colour scale is set to 25, as for the real data.

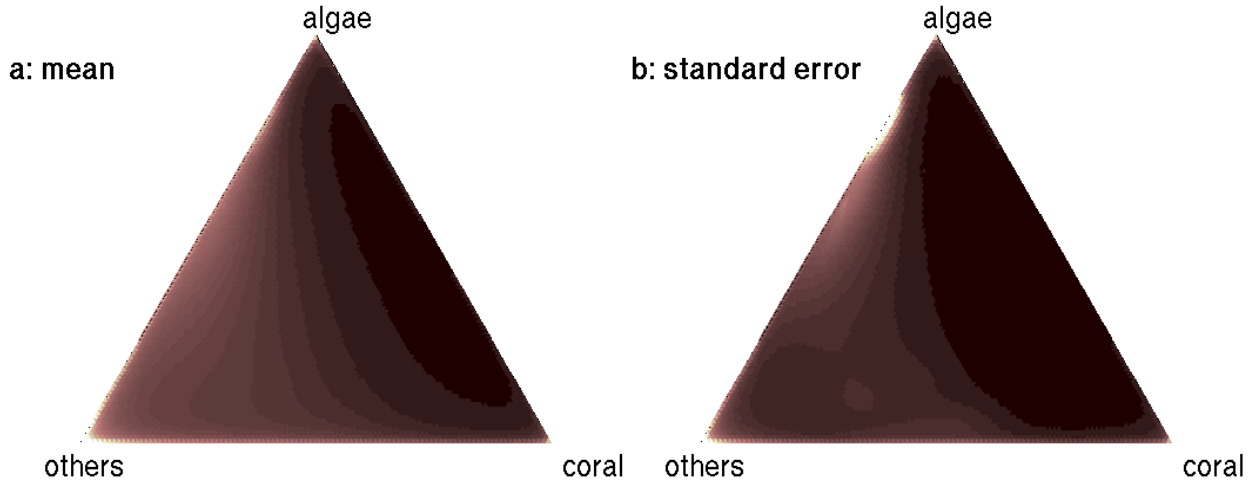

Figure S28: Mean (a) and standard error (b) of the stationary distributions of 100 replicates simulated under the model used to produce Figures S4 and S5, with multinomial measurement error, $m = 160$ points. Lighter colours are higher probability densities. The maximum of the colour scale is set to 25, as for the real data.

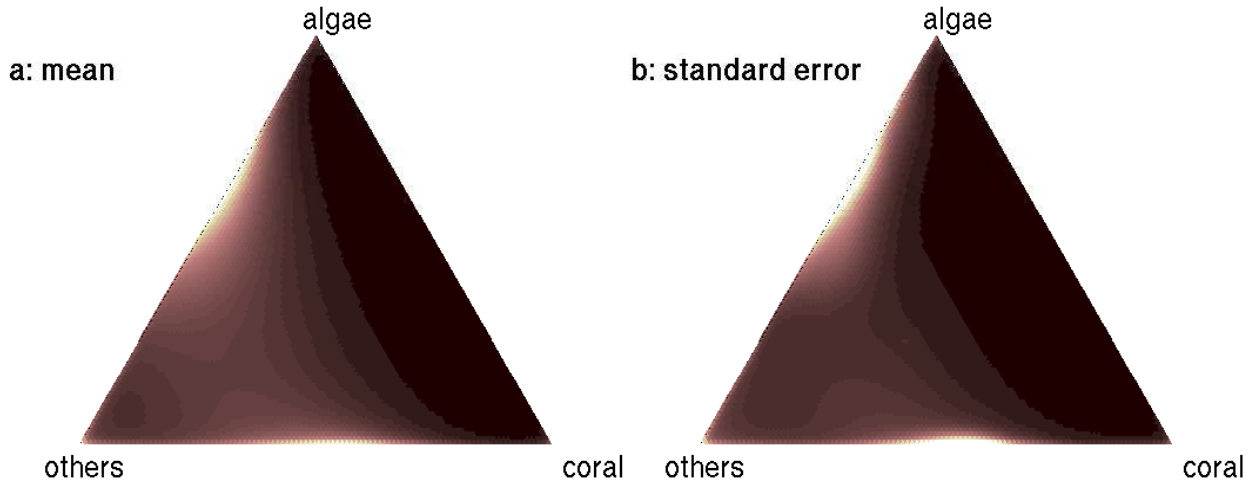

Figure S29: Mean (a) and standard error (b) of the stationary distributions of 100 replicates simulated under the model used to produce Figures S4 and S5, with multinomial measurement error, $m = 1000$ points. Lighter colours are higher probability densities. The maximum of the colour scale is set to 25, as for the real data.

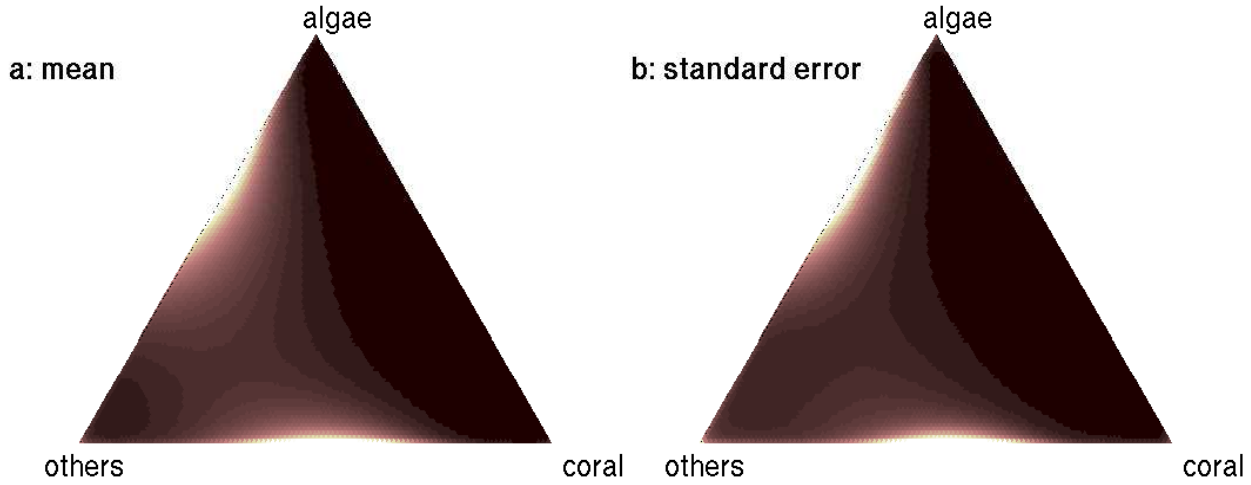

Figure S30: Mean (a) and standard error (b) of the stationary distributions of 100 replicates simulated under the model used to produce Figures S4 and S5, with multinomial measurement error, $m = 5000$ points. Lighter colours are higher probability densities. The maximum of the colour scale is set to 25, as for the real data.

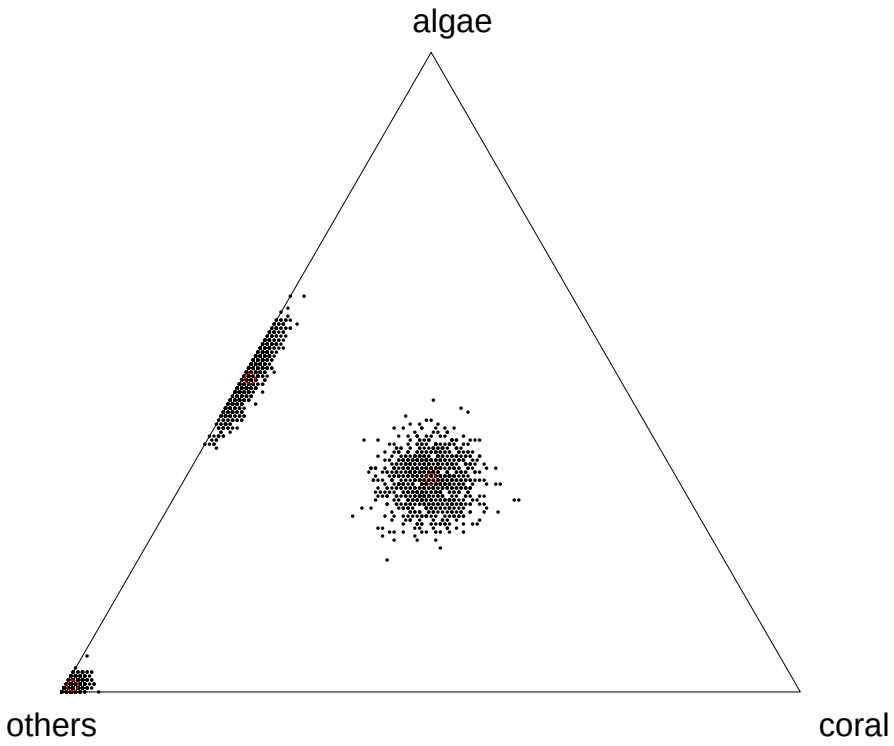

Figure S31: 1000 samples (black dots) of 160 points each from multinomial distributions with true states (red circles) $\mathbf{x} = [1/3, 1/3, 1/3]$ (centre), $[0.01, 0.49, 0.50]$ (upper left), and $[0.01, 0.01, 0.98]$ (lower left).

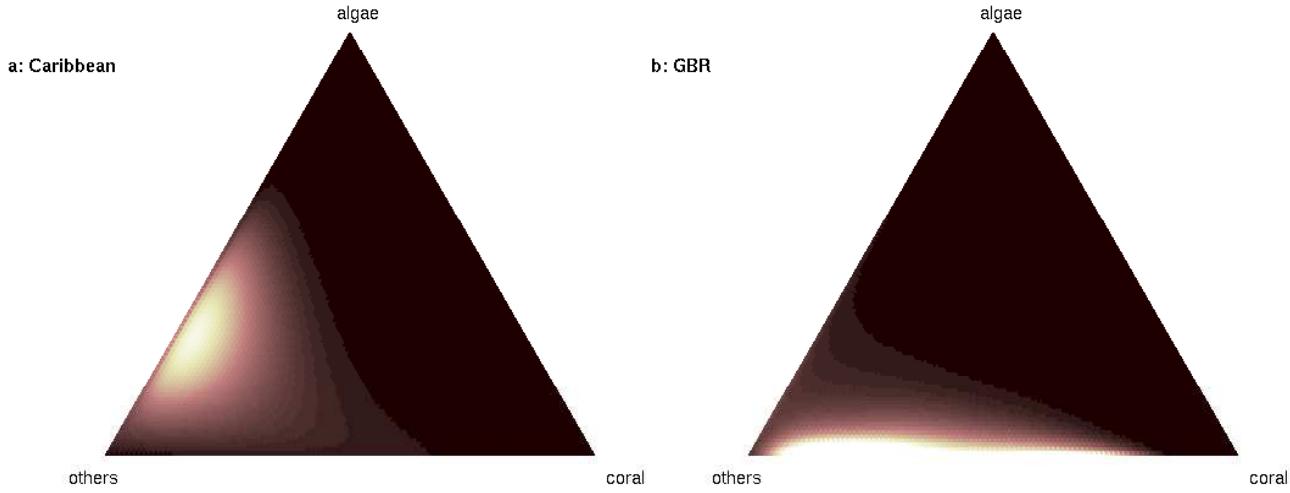

Figure S32: Estimated stationary distributions for (a) Caribbean and (b) Great Barrier Reef, with Reef Check data omitted. Lighter colours are higher probability densities. The maximum of the colour scale is set to 25.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.

Aitchison, J. (1989). Measures of location for compositional data sets. *Mathematical Geology*, 21:787–790.

Aitchison, J. (1992). On criteria for measures of compositional differences. *Mathematical Geology*, 24(4):365–380.

Aitchison, J. and Egozcue, J. J. (2005). Compositional data analysis: where are we and where should we be heading? *Mathematical Geology*, 37(7):829–850.

Aronson, R. B., Edmunds, P. J., Precht, W. F., Swanson, D. W., and Levitan, D. R. (1994). Large-scale, long-term monitoring of Caribbean coral reefs: simple, quick, inexpensive techniques. *Atoll Research Bulletin*, 421:1–19.

Bacon-Shone, J. (2003). Modelling structural zeros in compositional data. *Workshop - CoDaWork '03, Proceedings: Universitat de Girona, CD-ROM*. Available from: `http://ima.udg.es/Activitats/CoDaWork03/`.

Baker, C. T. H. (1977). *The Numerical Treatment of Integral Equations*. Clarendon Press, Oxford.

Berkelmans, R., De'ath, G., Kininmonth, S., and Skirving, W. J. (2004). A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef: spatial correlation, patterns, and predictions. *Coral Reefs*, 23:74–83.

Blackwood, J. C., Hastings, A., and Mumby, P. J. (2011). A model-based approach to determine the long term effects of multiple interacting stressors on coral reefs. *Ecological Applications*, in press.

Bruno, J. F. and Selig, E. R. (2007). Regional decline of coral cover in the Indo-Pacific: timing, extent, and subregional comparisons. *PLoS One*, 2:e711.

Bruno, J. F., Sweatman, H., Precht, W. F., Selig, E. R., and Schutte, V. G. W. (2009). Assessing evidence of phase shifts from coral to macroalgal dominance on coral reefs. *Ecology*, 90(6):1478–1484.

Bythell, J., Hillis-Starr, Z., and Rogers, C. (2000). Local variability but landscape stability in coral reef communities following repeated hurricane impacts. *Marine Ecology Progress Series*, 204:93–100.

Cheal, A. J., Delean, S., Sweatman, H., and Thompson, A. A. (2007). Spatial synchrony in coral reef fish populations and the influence of climate. *Ecology*, 88(1):158–169.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Connell, J. H. (1997). Disturbance and recovery of coral assemblages. *Coral Reefs*, 16:S101–S113.

Côté, I. M., Gill, J. A., Gardner, T. A., and Watkinson, A. R. (2005). Measuring coral reef decline through meta-analyses. *Philosophical Transactions of the Royal Society Series B*, 360:385–395.

Doucet, A., deFreitas, N., and Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In Doucet, A., deFreitas, N., and Gordon, N., editors, *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, New York.

Edmunds, P. J. and Bruno, J. F. (1996). The importance of sampling scale in ecology: kilometer-wide variation in coral reef communities. *Marine Ecology Progress Series*, 143:165–171.

Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *Annals of Statistics*, 9(3):586–596.

Ellner, S. P. and Rees, M. (2006). Integral projection models for species with complex demography. *American Naturalist*, 167:410–428.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

Gueorguieva, R., Rosenheck, R., and Zelterman, D. (2008). Dirichlet component regression and its applications to psychiatric data. *Computational Statistics and Data Analysis*, 52:5344–5355.

Hallin, M., Lu, Z. D., and Tran, L. T. (2004). Local linear spatial regression. *Annals of Statistics*, 32:2469–2500.

Harding, S., Lowery, C., and Oakley, S. (2000). Comparison between complex and simple reef survey techniques using volunteers: is the effort justified? In Moosa, M. K., Soemodihardjo, S., Soegiarto, A., Romimohtarto, K., Nontji, A., Soekarno, and Suharsono, editors, *Proceedings of the Ninth International Coral Reef Symposium, Bali, 23-27 Oct. 2000*, volume 2, pages 883–890.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43:525–546.

Hughes, T. P., Baird, A. H., Dinsdale, E. A., Moltschaniwskyj, N. A., Pratchett, M. S., Tanner, J. E., and Willis, B. L. (1999). Patterns of recruitment and abundance of corals along the Great Barrier Reef. *Nature*, 397:59–63.

Leujak, W. and Ormond, R. F. G. (2007). Comparative accuracy and efficiency of six coral community survey methods. *Journal of Experimental Marine Biology and Ecology*, 351:168–187.

Madin, J. S. and Connolly, S. R. (2006). Ecological consequences of major hydrodynamic disturbances on coral reefs. *Nature*, 444:477–480.

Masry, E. and Fan, J. Q. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, 24:165–179.

McClanahan, T. R. (2008). Response of the coral reef benthos and herbivory to fishery closure management and the 1998 ENSO disturbance. *Oecologia*, 155:169–177.

Miller, W. E. (2002). Revisiting the geometry of a ternary diagram with the half-taxi metric. *Mathematical Geology*, 34(3):275–290.

Mumby, P. J., Hastings, A., and Edwards, H. J. (2007). Thresholds and the resilience of Caribbean coral reefs. *Nature*, 450:98–101.

Murdoch, T. J. T. and Aronson, R. B. (1999). Scale-dependent spatial variability of coral assemblages along the Florida Reef Tract. *Coral Reefs*, 18:341–351.

Nadon, M.-O. and Stirling, G. (2006). Field and simulation analyses of visual methods for sampling coral cover. *Coral Reefs*, 25:177–185.

Ninio, R., Meekan, M., Done, T., and Sweatman, H. (2000). Temporal patterns in coral assemblages on the Great Barrier Reef from local to large spatial scales. *Marine Ecology Progress Series*, 194:65–74.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer, New York.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge University Press, Cambridge, England, second edition.

Rogers, C. S. (1993). Hurricanes and coral reefs: the intermediate disturbance hypothesis revisited. *Coral Reefs*, 12:127–137.

Steneck, R. S. (1988). Herbivory on coral reefs: a synthesis. In Choat, J. H., Barnes, D., Borowitzka, M. A., Coll, J. C., Davies, P. J., Flood, P., Hatcher, B. G., Hopley, D., Hutchings, P. A., Kinsey, D., Orme, G. R., Pichon, M., Sale, P. F., Sammarco, P., Wallace, C. C., Wilkinson, C., Wolanski, E., and Bellwood, O., editors, *Proceedings of the 6th International Coral Reef Symposium*, pages 37–49, Townsville, Australia.

Tanner, J. E., Hughes, T. P., and Connell, J. H. (1996). The role of history in community dynamics: a modelling approach. *Ecology*, 77(1):108–117.

van Woesik, R., Tomascik, T., and Blake, S. (1999). Coral assemblages and physico-chemical characteristics of the Whitsunday Islands: evidence of recent community changes. *Marine and Freshwater Research*, 50:427–440.

von Winckel, G. (2005). n-dimensional simplex quadrature. Available from: `http://www.mathworks.com/matlabcentral/fileexchange/9435-n-dimensional-simplex-quadrature`.

Wooldridge, S., Done, T., Berkelmans, R., Jones, R., and Marshall, P. (2005). Precursors for resilience in coral communities in a warming climate: a belief network approach. *Marine Ecology Progress Series*, 295:157–169.