Measuring the benefits of protecting reefs from pollution using models of species turn-over along gradients of threat

Estimating the footprint of pollution on coral reefs using models of species turn-over

Christopher J. Brown ^1 Richard Hamilton[1]

1. Australian Rivers Institute, Griffith University, 170 Kessels Road, Nathan, Queensland, 4111, Australia
2. The Nature Conservancy, Asia Pacific Resource Centre, 48 Montague Road, South Brisbane, QLD 4101, Australia
3. ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia

Intended for submiossion to: Conservation Biology

Draft compiled on 8th June 2017

Abstract

Ecological communities typically change along gradients of human threat, though it is difficult to estimate the extent of impacts for diffuse threats like pollution. Here we develop a joint model of benthic habitats on lagoonal coral reefs and use it to infer change in benthic composition along a gradient of distance from logging operations. We detect compositional turnover across the gradient and use the model to predict the footprint of turbidity impacts from logging. We then apply the model to predict impacts of recent logging activities, finding recent impacts to be small, because recent logging has occurred far from lagoonal reefs. Our model can be used more generally to estimate the extent of impacts from diffuse threats to ecosystems and evaluate the benefits of conservation actions for ecosystems.

Keywords

Bayesian statistics, ecological indicator, threat mapping, coral reef, water quality, latent variable, pollution, logging, conservation evaluation, multi-species distribution modelling

Introduction

Diffuse threats, like pollution, can be a major cause of ecological change. Determining the extent and impact of diffuse threats is an important concern for ecological science and management. For instance, the extent of pollution impacts on marine benthic communities is often used to estimate the benefits of management actions aimed at its mitigation (Bryan 1971; Warwick and Clarke 1991; De'ath and Fabricius 2010). Evaluating the ecological benefits of mitigative efforts requires spatially comprehensive mapping of the ecosystems that are significantly affected. When the impact of a threat on different species varies, changes in community composition can be used to infer the spatial extent of a diffuse threat (Warwick 1993). For instance, coastal development can increase turbidity of waters, causing shifts in coral reef communities toward stress tolerant species (e.g. Guest et al. 2016). Thus, species turnover, or beta-diversity, can be used to determine how far along a gradient one must move before the community is significantly changed (Anderson et al. 2011; Socolar et al. 2016).

---

[1]2, 3

The spatial extent of diffuse impacts to communities can be challenging to infer from patchy measurements. Here we refer to 'community composition' as any multivariate measure of the organisms that make-up a community, including both species and higher taxonomic groupings and measurements made on counts of individuals, occurrences of a taxonomic group or per cent cover. We refer to these different measurement types collectively as 'abundances'. The analysis of community change over physical gradients presents many technical challenges. In particular, there are often many species sampled but few sites. Methods for the analysis of change in multivariate abundances across environmental gradients have undergone decades of development (e.g. Kruskal 1964; Anderson et al. 2011; P. Legendre and Legendre 2012). However, current methods are limited in that they are typically not based on models of abundances so, while they can be used to test for signficant differences in community composition, they cannot predict abundances to unsampled sites (Warton et al. 2015). Prediction at unsampled sites is crucial for estimating the extent of a diffuse threat.

Multivariate statistics is currently entering a new phase of development that allows statistical prediction of abundances for multiple organisms in a similar fashion to species distribution modelling (Warton et al. 2015). These new methods, termed 'joint modelling' build on generalised linear models to allow inclusion of multiple species and their interactions with each other and environmental covariates (Hui et al. 2015; Warton et al. 2015). The linear modelling framework also means that joint models can be used to estimate uncertainty in species-environment relationships. Direct inclusion of environmental covariates within models allowing predictions of abundances to unsampled sites (Hui et al. 2015). Abundance models also have increased power to detect change in abundances when compared to traditional ordination methods. Covariance in species' abundance can also be modelled with latent (unmeasured) variables, which allow users can make inferences on unmeasured environmental gradients and interactions among species (Warton et al. 2015).

Here we sought to develop a new type of joint model that could estimate the areal footprint of diffuse threats, like pollution, on ecological communities. We developed a model that builds on recent methods for Bayesian ordination (Hui et al. 2015; Hui 2016). The new development is the inclusion of constrained latent variables that represent turn-over in the composition of the benthic community across a gradient of pollution impacts. Our model is a generalisation that allows for covariates to effect the latent variables, rather than directly affecting species abundances. The model could also be viewed as a type of Bayesian structural equation model (e.g. Joseph, Preston, and Johnson 2016), or as a Bayesian analog to constrained factor analysis. By using distance from a source of pollution as the covariate the constrained latent variable can be interpreted as an unobservable gradient of community change. The rate of change in the latent variable in response to the threat represents community turn-over and so is interpretable as a measure of beta-diversity (Anderson et al. 2011).

To illustrate the utility of joint modelling in conservation, we apply the model to survey data of coral reef communities in the Kia region of Solomon Islands. Around Kia sediment run-off from logging operations has degraded lagoonal reefs (Hamilton et al. 2017). We applied the joint model to surveys of lagoonal reef communities that were conducted across a gradient of distances from logging operations. We aimed to address three objectives in the case-study. First, we sought to quantify how different components of the benthic community respond to sediment run-off from logging and identify interactions among benthic habitats. We also asked whether these community responses were consistent with studies of coral reef communities from other regions. Second, we aimed to estimate the areal footprint of logging impacts on benthic communities. Finally, new illegal logging has occurred in the region since the original surveys and here we estimate how much more reef has been affected by this new logging. More generally, our objective is to use the case-study to illustrate how joint modelling can be applied to estimate the impact of new developments, or the benefits of conservation actions.

Methods

First we describe a general framework for modelling community composition as a function of latent variables and their covariates. Then we describe a specific application of the model to benthic communities in the Solomon Islands. Finally, we test the model's ability to make correct inferences on simulated data.

Model framework

We modelled abundances of multiple response groups as a function of multiple latent variables (Fig 1). The method built on existing models for Bayesian ordination (Hui 2016). The response groups could represent species, taxonomic groups, or as in our case-study below, habitat categories. The abundance of each group was modelled as a function of a linear predictor:

$$\eta_{i,j} = \theta_i + \alpha_j + \sum_{k=1}^{K} \beta_{j,k}\gamma_{i,k}$$

Equation 1

Where θ_i was an offset term that accounted for variation in sampling intensity across sites, $\eta_{i,j}$ was the value of the linear predictor for group j at site i, α_j is a group level intercept term, which allows among group variability in average abundances across all sites. The summation is over K latent variables which take values $\gamma_{i,k}$ at each site. Finally, $\beta_{j,k}$ is a group's loading on a given latent variable.

The linear predictor $\eta_{i,j}$ could be used within a range of distributions for abundance such as the poisson or bernoulli distributions for count and occurrence data respectively. The model could also account for over-dispersion, for instance, through use of a negative-binomial distribution, however we do not describe that extension here. Below for the case-study we describe application to per cent cover data, so we use the linear predictor within a multinomial model.
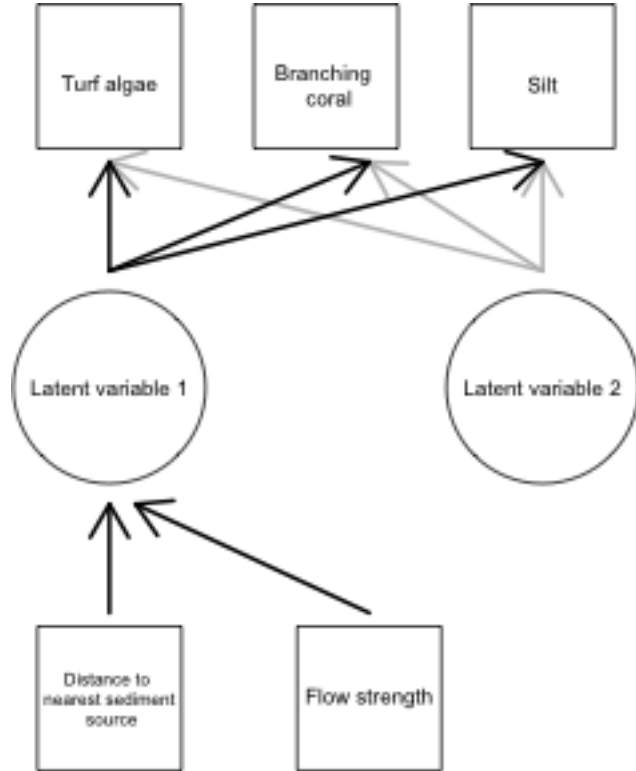

Figure 1 Directed graph giving an example for the structure of our Bayesian latent variable model applied to coral reef habitats in Solomon Islands. Squares indicate measured variables and circles indicate latent variables. Arrows indicate model effects, with gray and black arrows indicating the effects relating to different latent variables. ***NB need to add errors to figure. . . .***

The latent variables were sampled from normal distributions:

$\gamma_{i,k} \sim N(\kappa_i X, 1)$ *Equation 2*

Where there are K latent variables across the i sites, κ_k is a vector of coefficients and X is a matrix of covariate values across the sites. Latent variables for residual correlations have $\kappa_k = 0$. Thus, the key modification in this model from past latent variable models is that we allow for the possibility of constraining some latent variables by covariates.

An alternative model would be to specify direct effects of the covariates on group's abundances [hui2015model]. The primary difference between a direct effects model and our proposed model is that our proposed model estimates only one coefficient for the spatial attenuation of the threat, whereas the direct model estimates one attenuation coefficient per group. This difference results in a subtle but important difference in interpretation of environmental effects to abundances. The coefficients for the direct model are interpretable in terms of change in abundance of individual groups with environmental change, for instance, a poisson model of abundance would estimate the rate of change in the log of abundance over environmental change. Our proposed approach loses some interpretability at the individual group level, but gains interpretability at the community level. The attenuation coefficient now represents the rate of change in an ordination of community across the environmental gradient.

The constrained latent variable estimated in our proposed approach could be interpreted in multiple ways. If abundances are taken to be imperfect indicators of a latent environmental state, like turbidity, then the latent variable represents our best estimate of for the dispersion of turbid waters. Taking the latent variable as a measure of turbidity also means we should interpret its coefficients ($\kappa_k = 0$) as representing attentuation of turbidity over space. Abundances could also be interpreted as imperfect indicators of an underlying community gradient, in which case the latent variable represents our best estimate of the community's state with respect to the gradient. If we take the community-centric interpretation then the latent variable's coefficient is a measure of community turn-over across the gradient (i.e. beta-diversity) (Anderson et al. 2011).

Priors for species loadings ($\beta_{j,k}$) and covariate effects ($\kappa_k = 0$) are specified using normal distributions. We use vaguely priors for covariate effects (mean = 0, variance = XXX) and vaguely informative priors for species loadings (mean = 0, variance = 20) to aid convergence (Hui 2016). Vaguely informative priors are appropriate here because we standardised data, so we do not expect variances or effects to be $>> 1$.

Case-study

We applied joint modelling to change in cover of benthic habitats across a gradient of sedimentation in the Kia District of Isabel Province, Solomon Islands. Logging in Kia district has removed 50% of forests over the last 15 years. Typical operations haul logs to log ponds, locations on the coast where mangroves are bulldozed so logs can be stored then transported elsewhere. Log ponds result in significant sedimentation entering nearshore environments, which smothers corals and causes declines in branching corals and fish that depend on those corals for habitat (Hamilton et al. 2017). The southern part of the district has extensive logging (Hamilton et al. 2017), whereas the northern section was unlogged when our survey data were collected in 2013.

Previously we estimated region-wide loss of fish habitat based solely on the area of reefs within each region (Hamilton et al. 2017). Here we seek more accurate estimates for the extent of degradation of reef habitats and also to investigate relationships among different habitat types and sediment. Additionally, some illegal logging has occurred in the north-western part of the region, which before 2013 was unlogged. The recent illegal logging may have affected the relatively pristine northern reefs, which are important habitat for bumphead parrotfish (*Bolbometopon muricatum*), a locally important fishery species that is also listed as threatened on the IUCN Red List(Hamilton et al. 2017). Thus a second objective of this study was to estimate the area of lagoonal reefs that may have been affected by this illegal logging.

The data include surveys of benthic communities at 49 sites, conducted in October 2013 (Hamilton et al. 2017). Sites were surveyed using a modified version of the point intercept method (J. Hill and Wilkinson 2004). Benthic habitats were recorded every 2 m along a 50 m transect, taking three measurements at each point, which were located directly below and one metre to the left and right of the transect. There were five transects of 75 points at each site, making a total of 375 points per site. Benthic habitats were categorised as

per English et al. (1994), however we added categories for Acropora Branching Dead and Coral Branching Dead. Water clarity was also measured in deep water near to survey sites using a Secchi disc.

For analysis we aggregated categories into 17 focal groups. Aggregation sped computations and emphasised change in the habitat types we were most interested in. We also aggregated categories to avoid having numerous infrequently recorded categories that had many zeros. Rare categories would require additional modifications of the model to allow for over-dispersion (e.g. zero-inflated models Martin et al. (2005)), which is beyond the scope of this initial analysis.

Cover (number of points out of 375) of the benthic habitats at each site was modelled as a multinomial variate, using a mix of poisson distributions (Gelman and Hill 2006). The poisson model for the count of points belonging to each habitat was specified using a log-link as per equation 1. To ensure the model parameters are identifiable the habitat-level intercepts (α_j) are modelled as a random effect (Gelman and Hill 2006). Predictions for proportional cover of each habitat at each site can then be obtained by normalising predictions for the count of each habitat by the sum of expected counts across all habitats:

$$p_{i,j} = \frac{e^{\pi_{i,j}}}{\sum_{j=1}^{J} e^{\pi_{i,j}}}$$

Equation 3

Where $\pi_{i,j}$ is the expectation for the count of habitat j at site i, α_j is a habitat-specific random effect, β_j.

We addressed the first objective, which was to quantify community responses to logging, by performing model selection to identify and best model and then visualising mean responses to log ponds (with 95% credibility intervals) for the entire community and each component habitat.

Our first latent variable was modelled as a function of distance to log ponds and flow strength. We fitted two alternate models, one where the distance covariate was the distance to the nearest log pond. A second cumulative effects model had distance as the sum of distances to all log ponds. For both models we standardised the distance metric by dividing by its standard deviation. We then compared the nearest and cumulative distance models using the Watanabe-Akaike information criterion (WAIC) (Vehtari, Gelman, and Gabry 2016) to see which had a superior fit. Flow was classified as either low or high, based on experience of the survey divers. Previous analysis indicated that flow strength was important for explaining cover of branching corals (Hamilton et al. 2017). For all covariates we estimated mean effects with 95% credibility intervals, where a 95% CI that does not overlap zero is statistically significant. As an additional test of the model we compared the estimates of the sedimentation latent variable to Secchi depths that were measured during surveys (Hamilton et al. 2017).

We also included two unconstrained latent variables to account for residual correlations among habitat categories. We decided to use two latent variables, because results from two latent variables can be easily visualsed (Hui 2016). IN additional simulations we refit the models with up to 6 unconstrained latent variables. We also fit a model with two unconstrained latent variables and no constrained latent variables as per (2016) and compared its fit using the WAIC to the nearest and cumulative distance models.

We checked model fits by comparing predicted to observed per cent covers for each habitat type and also checked counts of points using rootograms (Kleiber and Zeileis 2016). We calculated the bias and variance in predicted to observed cover levels.

We addressed our second objective by predicting the constrained latent variable across the entire study region, using as a covariate the over water distance to the nearest log pond. Because we did not have flow estimates for the entire study region, we assumed low flow at all sites, which was the dominant condition at surveyed reefs (which had the greatest sedimentation impacts). We used the predictions for the unconstrained latent variable to map the sedimentation gradient across unsampled lagoonal reefs. Because the latent variable is unitless, we mapped its values as the probability a given reef would have a value of the latent variable equal to, or worse than, the average value for the nearest reef to the log pond. We calculate this statistic as follows:

$$pr(L_d \leq \widehat{L_0}) = CDF(\widehat{L_0}, \mu = \dot{\kappa}D, \sigma = 1)$$

5

Equation 4

Where L_i is the value of the constrained latent variable at distance d, $\widehat{L_0}$ is the average value of the condition variable at the site that is nearest to any log pond, CDF is the cumulative distribution function for the normal distribution and $\acute{\kappa}$ is an estimate for the effect of distance on the condition variable. We chose to use the modal estimate for $\acute{\kappa}$. The standard deviation was fixed at 1, as per *Equation 2*. Thus, the maximum probability was 50%, indicating a reef that was the same as the closest site. Lower values indicate the reef has a lower probability of being in worse condition than the average closest site. For reefs that were outside the range of distances from log ponds observed in the surveys we used the minimum and maximum observed distances to predict the condition variable, which avoided extrapolation.

Our final objective was to estimate the area of lagoonal reefs affected by new logging operations. We repeated the calculations for the area of reef affected by logging with new logging that has recently occurred logging on the north-western Coast (three log ponds at: [7.43639^oS, 158.19973^oE], [7.42830^oS, 158.20321^oE] and [7.42558^oS, 158.20111^oE]) and then we calculated the area of reefs that are predicted to be affected by increased turbidity.

We used JAGS (Plummer and others 2003) for parameter estimation, run from the R programming environment (R Core Team 2016) using the package rjags (Plummer 2016). For each model we ran a single chain with 100 000 samples, thinning for every 10 sample. We used only a single chain to avoid issues with parameter switching on the latent variables (Hui 2016). We checked the Hellinger distance statistic (Boone, Merrick, and Krachey 2014) over multiple runs with single chains to confirm model convergence.

Results

The effect of cumulative distance to log ponds on the constrained latent variable was weaker (Fig 2, 95% C.I.s overlap zero), but tended in the same direction as high flow conditions. The effect of minimum distance was also in the same direction as high flow conditions and the lower 95% CI was slightly greater than zero. The mean effect of high flow was ~4 times greater than the mean effect of minimum distance on the latent variable. Thus, we can interpret a change from low to high flow as having an equivalent effect on community structure as moving 2.46 km further away from log-ponds.

The cumulative and nearest distance models had similar WAIC values, suggesting both are equally plausible explanations of gradients in the community (Table 2). We proceed with further analysis using the nearest distance model, because interpretation for this model is more straightforward than the cumulative distance model.
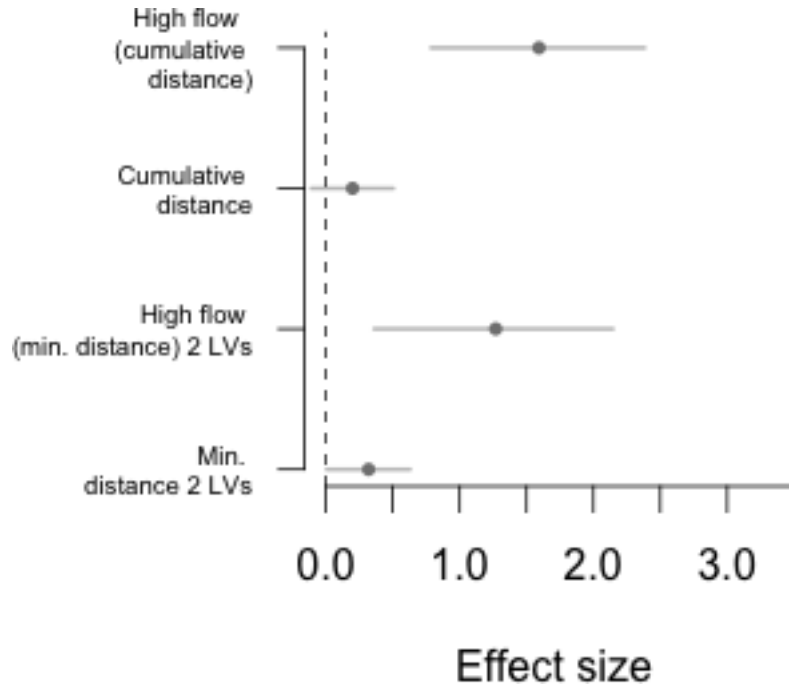
Figure 2 CIs on distance (continuous) and high flow (binary) effects for nearest and cumulative distance models. Bars indicate 95% credibility intervals and points mean effects estimates. Cumulative and minimum distances were normalised so an effect size of 0.04 (minimum distance model) represents change in the value of the latent variable by 1 SD (= 614 metres).

Table 2 Model comparison using the WAIC. LV: number of unconstrained latent variables.

Model	WAIC
Bayesian ordination, 2 LVs	15822
Cumulative distance, 2 LVs	12616
Nearest distance, 2 LV	12897

Simulations with increasing number of LVs - how much variance is explained by constrained LV?

We checked the nearest distance model for bias in its predictions of each habitat separately (add supplement table of bias and r^2 estimates). In general the model fitted most habitats with low bias. Bias was higher for rare habitats, which had many zero observations.

The loadings of the habitats on the constrained latent variable indicated that it represents a gradient from high cover of branching Acropora, branching corals, dead branching corals, other habitats and rare corals far from log-ponds, towards higher cover of sand, macro-algae and turf algae closer to log-ponds (figs 3a, 4). At sites far from log ponds, branching *Acropora* were commonly observed growing out of sand, so the high cover of sand, but low cover of *Acropora* at sites near to log ponds ostensibly indicates the loss of *Acropora*. Massive corals and dead corals were also associated with sites closer to log ponds (Fig 3a), however their overall change increase closer to log ponds was small (Fig 4), suggesting they maintained their cover across the gradient, rather than replacing other habitats. Overall, cover of algae was low (<10%) si so that sites near to log ponds were dominated by sand and massive corals (Fig 4). Thus, the constrained latent variable represented a gradient from communities dominated by stress tolerant corals and unconsolidated sediments toward communities with high cover of structurally complex corals.

The unconstrained latent variables indicated positive associations among cover submassive corals, tabulate Acropora, branching Acropora and other corals (Fig 3). These habitats were negatively correlated with

7

macroalgae, halimeda and possibly rock (Fig 3b, 95% CI overlaps with zero). Algal assemblages and rock were also positively associated and traded-off against halimeda (Fig 3c).
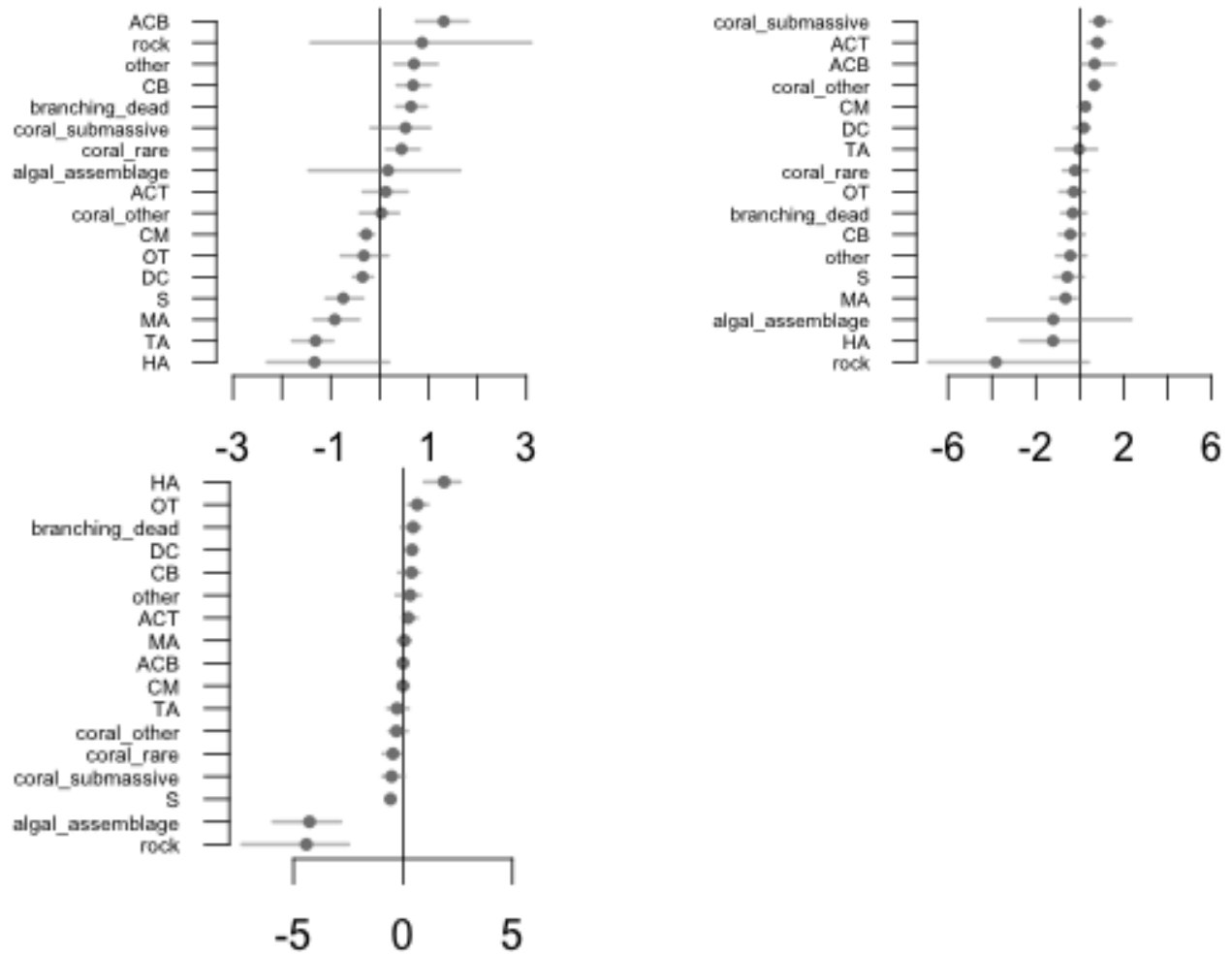

Figure 3 Mean estimates and CIs for loadings of each habitat category on the (a) constrained latent variable and (b-c) two unconstrained latent variables. Loading signs on the constrained latent variable (a) are fixed so that positive values indicate a habitat's cover increases further from log-ponds.
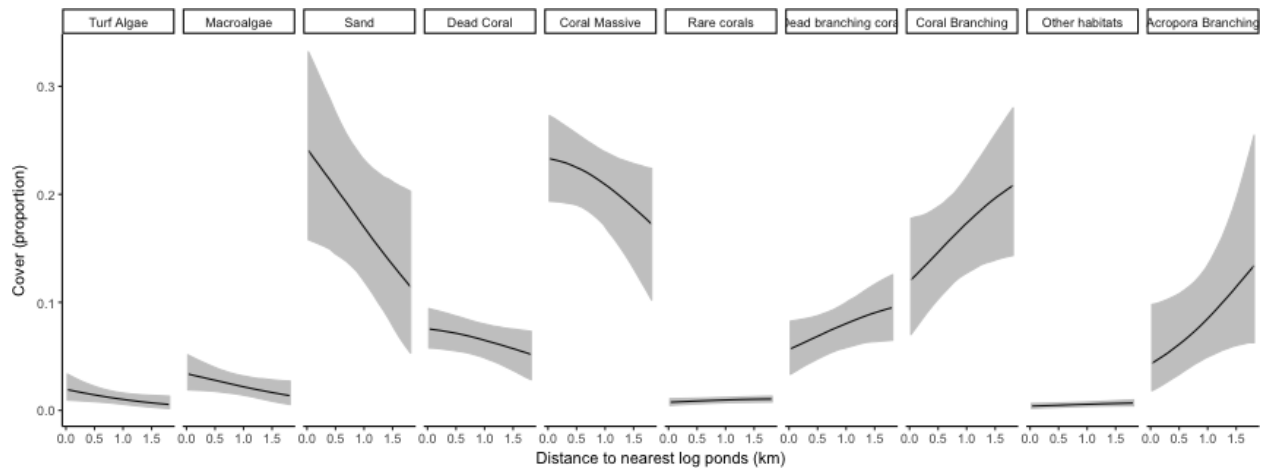

Figure 4 Cover against distance to nearest log-ponds for habitats with a significant response to the latent

8

variable for water quality. Significant was defined as 95% CIs not overlapping zero.

Mean estimates of the latent variable were negatively correlated with Secchi depth ($\rho = XXX$), where lower values of the latent variable indicate benthos associated with more turbid waters. The relatively weak correlation may be due to high temporal variability in Secchi depths that were measured only once at each site and the displacement of Secchi measurements from survey sites to deep channels.

By predicting the values of the condition variable across all reefs, we can estimate footprint of logging impacts on reefs, as the probability a reef at a given distance will have worse condition than the reef nearest to any log pond. We estimated that 1% (0 - 55 95% C.I.) of reefs had a 50% probability, 56% (0 - 59 95% C.I.) of reefs had a 40% probability and 94% (100 - 68 95% C.I.) of reefs had a 25% probability. Overall, most lagoonal reefs on the southern side of have been affected by log ponds with >40% probability (Fig 5), however the broad 95% C.I. bounds indicate considerable uncertainty over the status of individual reefs.

We also estimated the area of reef that has been protected by actions to stop logging on the north-western side of Kia. We predicted with the inclusion of new log ponds affected 1% (0 - 56 95% C.I.) of reefs with 50% probability or 61% (0 - 64 95% C.I.) of reefs with 40% probability, amounting to an extra 0 to 5 % of reefs or 0 to 148 hectares of reefs.

Figure 5 Map of study region showing log ponds, survey sites and probabilities that a reef would have worse condition than the average reef that is closest to any log pond. The gradient shows the spatial field for probability and takes its maximum value at 50%. For instance, a probability of 60% indicates that reefs in this location would have a 60% probability of having worse condition than the average reef that is close to a log pond. The probabilities are predicted to all ocean areas (not just lagoonal reefs) for visualisation purposes. Note that uncertainty on the probability field is broad, but but is not shown here for visualisation purposes.

Discussion

We used joint modelling to estimate the ecological impacts of logging on coral reef communities. We found there was a gradient in reef communities with reefs closer to log ponds lacking complex coral habitats and instead having higher cover of massive corals and unconsolidated sediment. This gradient is consistent with studies of turbidity gradients from other regions of the Pacific. For instance massive corals are highly tolerant of low light conditions and high sediment loads (ref). Reefs in Palau and Fiji have also been observed to lack cover of structurally complex hard corals, like branching *Acropora* closer to turbid rivers (Golbuu et al. 2008, Brown et al. (in review)).

The joint modelling used here enables prediction of abundances of community members at unmeasured sites with uncertainty bounds (Warton et al. 2015). We estimated community structure across all of Kia's lagoonal reefs and found that ~50% of the region's reef communities have been affected by existing log-ponds with 40% probability, though uncertainty bounds on this estimate were broad. Lagoonal reefs with structurally complex corals are nursery habitat for locally important fishery species, so community change across these reefs has likely affected recruitment of juveniles to the fishery and local livelihoods (Hamilton et al. 2017). While uncertainty around the extent of impacts from an individual log pond was high, we can be confident that most southern reefs are affected by log ponds because logging in the southern region has been extensive. Globally, there is concern that loss of structurally complex habitats will cause widespread declines in reef fish species and affect human communities that depend on reef fisheries (ref). Pollution, in particular low water clarity from sediment run-off is a key contributor to the decline of structurally complex reefs (ref). Our joint modelling approach thus provides an important tool for assessing the footprint of human activities that have increased turbidity and sedimentation on reefs.

Predictions of abundance are useful for conservation planning because the benefits of conservation actions and the impacts of development on abundance can be quantified. We predicted that recent illegal logging on the north coast of Kia may have further degraded only a small area of lagoonal reefs (<5%). This finding was consistent with our intuition, because the new logging has occurred on exposed coastline with very little lagoonal reef nearby. However, the broad width of the uncertainty limits indicates more work is needed to
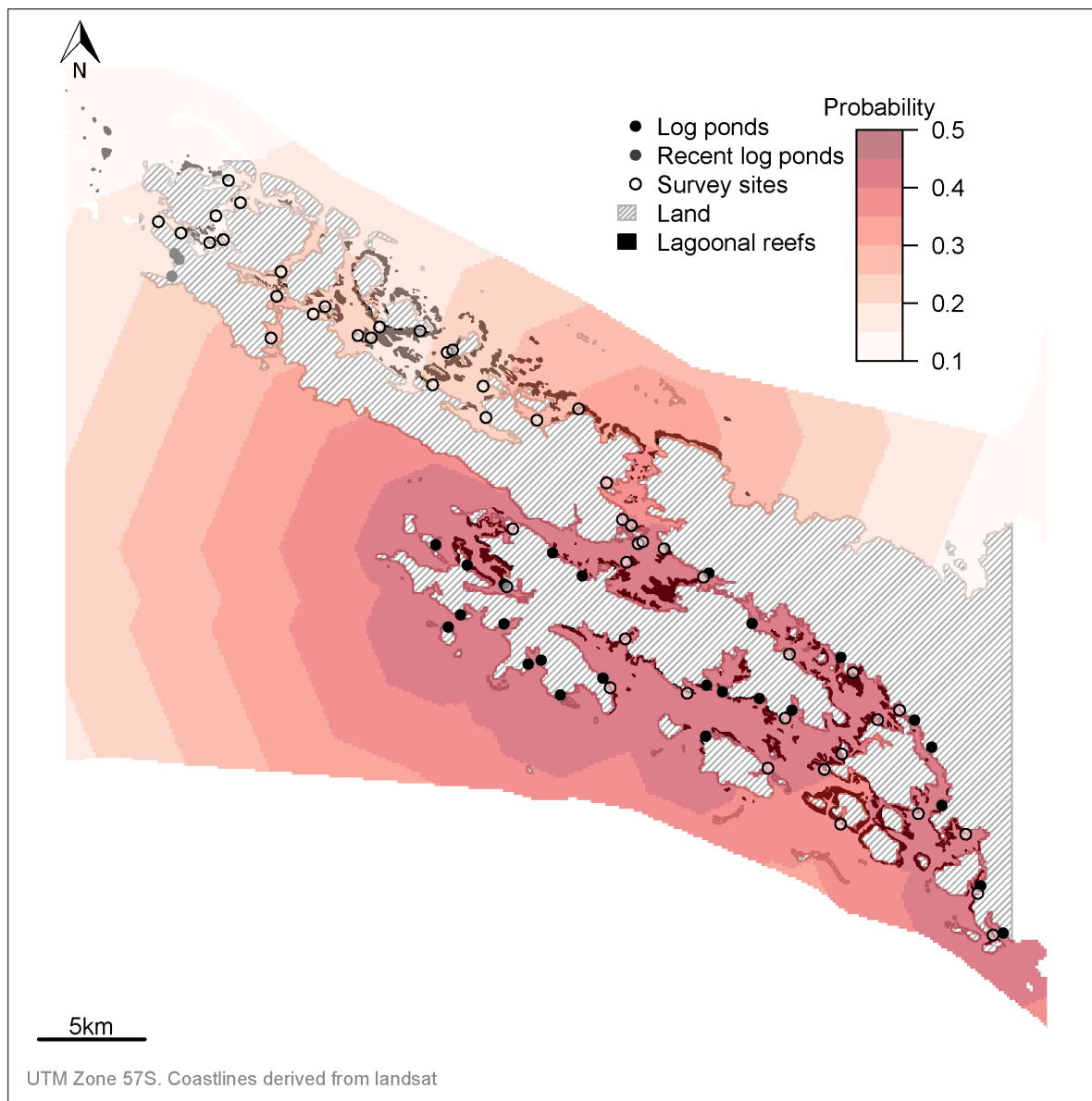
Figure 1:

more accurately quantify impacts of individual log-ponds. Thus, the joint modelling method can be used to evaluate the impacts of development, but may also indicate that further research into impacts is warranted.

Our analysis of community turnover across the pollution gradient has several advantages over the traditional approach of using distance-based multivariate statistics. We modelled abundance (habitat cover) directly, so the model could be used to make direct predictions of abundance across the pollution gradient, whereas distanced-based methods cannot predict changes in abundance directly. Second, we could make direct predictions for the constrained latent variable that represented community state. Thus, we could predict the areal extent of pollution impacts and the area of new reefs affected by recent logging activities. Such predictions are useful for evaluating the impacts of development actions and can be used to bolster support for further conservation (e.g. refs?).

A second benefit of our approach over distance-based methods is that we could include multiple latent variables that can help explain associations among benthic groups that were not accounted for by the environmental gradient (Warton et al. 2015). Associations present in the unconstrained latent variables can be used to infer on ecological interactions (Warton et al. 2015). For instance, we found that many coral groups were negatively associated with macro-algae, which may reflect competition and interference between these groups, such as toxins secreted by macro-algae that prevent coral establishment (Hay ref...). Modelling of species interactions is an important future task for multi-species distribution modelling, though species interactions are often not apparent in pure abundane data (ref). Future research should explore using prior information on species interactions within joint models. For instance, we could plausibly include a negative affect of macro-algae on corals a-priori in our model by biasing the coefficients for these groups in opposite directions on one of the latent variables. Further analyses should explore the utility of prior information to improve the accuracy of model predictions.

Generalised linear models provide flexible tools to model mean-variance relationships and different types of response data. For instance, we were able to model data on % cover, and % cover data often has the highest variance at intermediate values. Appropriate choice of mean-variance relationships can enhance the power of model-based methods to detect change in less common species when compared to distance-based methods (Hui et al. (2015)). An alternative approach would be to use structural equation modelling or factor analysis, indeed our model is a form of Bayesian factor analysis. An advantage of our model is that it allows for different types of error distribution, so can be applied to non-Gaussian data like counts, % cover or species occupancy rates.

Formalsing/predicting role of beta in conservation with models socolar2016should

The estimated rate of change in the constrained latent variable can be interpted in two ways. First, it represents a measure of directional turn-over over an environmental gradient and can thus be interpeted as a kind of beta-diversity (Anderson et al. 2011). If similar models were fit for different regions, or a single model for multiple gradients, then we could compare beta-diversity across those gradients or regions.

The estimated rate of change could also be interpreted as the rate of attentuation in turbidity, if we take benthic communities as indicators of environmental change. It would be fruitful in future research to compare these models to direct measurements of turbidity, or numerical models of turbidity and compare the spatial dispersion of the threat. Because communities integrate environmental change over time they may therefore offer a more rapid way to assess environmental change than measuring pollution direclty, which can be highly variable through time.

There are some technical challenges to further development of our models and to further application to evaluating conservation success. First, joint models currently require a higher degree of technical sophistication to implement than traditional distance based metrics, which many ecologists are familiar with. They also generally require greater computational power. Therefore, their application may not be straightforward in all contexts without specialised training. However the availability of programming packages like BORAL is increasing the accessibility of these techniques.

We also found that the amount of variance in community abundances explained by the constrained latent variable decreased as we increased the number of unconstrained latent variables. This occurred because the unconstrained variables have greater flexibility to explain the community abundances. Therefore, the analyst

needs to be careful to select the number of constrained and unconstrained variables. We suggest doing so a-priori, but further work is needed to provide more general advice on this issue.

A caveat is that model output can be sensitive to design. So analyst needs to think apiori how to build model.

All of the caveats normally associated with the analysis of spatial data also apply to our model. For instance, we did not account for spatial auto-correlation. In theory it would be possible to model spatial autocorrelation in the abundances directly or in the latent variable, though we are not aware of anyone who has attempted this for a multi-species distribution model.

When evaluating conservation success using distribution models we need to be careful of confounding covariates. For instance, if distance to log ponds was associated with a confounding, but unmeasured environmental covariate such as exposure to wind and waves, we could falsely attribute community change driven by exposure to turbidity. If this occurred, we would overestimate both the impact of existing log-ponds and the impact of future logging. We do not know of any important confounding gradients in Kia We do not know of any such important confounding gradients in this Kia study, since all sites surveyed were lagoonal fringing reefs that were sheltered from significant wave exposure. More generally, analysts should be careful to design surveys to avoid confounding gradients and should also include environmental variables that they have an a-priori reason to expect will be causal drivers of community change.

More general models like this and BORAL should be usueful for evaluating conservation actions. use boral if you want to predict species change directly. Use this if you want to predict univariate measure.

We estimated the impact of logging on coral reef communities. To do so, we developed a flexible joint modelling approach. This approach enables predictions of species turnover and predictions for the extent of impacts from diffuse threats. This will help conservation decision makers evaluate the cost of human threats to ecosystems and also the benefits of avoiding future threats.

Acknowledgements

References

Anderson, Marti J, Thomas O Crist, Jonathan M Chase, Mark Vellend, Brian D Inouye, Amy L Freestone, Nathan J Sanders, et al. 2011. "Navigating the Multiple Meanings of β Diversity: A Roadmap for the Practicing Ecologist." *Ecology Letters* 14 (1). Wiley Online Library: 19–28.

Boone, Edward L, Jason RW Merrick, and Matthew J Krachey. 2014. "A Hellinger Distance Approach to Mcmc Diagnostics." *Journal of Statistical Computation and Simulation* 84 (4). Taylor & Francis: 833–49.

Brown, CJ, Jupiter S, Lin H-Y, Albert S, Klein CJ, Maina JM, Tulloch V, Wenger A, and Mumby PJ. in review. "Habitat Change Mediates the Response of Coral Reef Fish Populations to Terrestrial Run-Off." *Marine Ecology Progress Series.* MEPS.

Bryan, GW. 1971. "The Effects of Heavy Metals (Other Than Mercury) on Marine and Estuarine Organisms." *Proceedings of the Royal Society of London B: Biological Sciences* 177 (1048). The Royal Society: 389–410.

De'ath, Glenn, and Katharina Fabricius. 2010. "Water Quality as a Regional Driver of Coral Biodiversity

and Macroalgae on the Great Barrier Reef." *Ecological Applications* 20 (3). Wiley Online Library: 840–50.

English, SS, CC Wilkinson, and VV Baker. 1994. "Survey Manual for Tropical Marine Resources." Townsville: Australian Institute of Marine Science.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge university press.

Golbuu, Yimnang, Katharina Fabricius, Steven Victor, and Robert H Richmond. 2008. "Gradients in Coral Reef Communities Exposed to Muddy River Discharge in Pohnpei, Micronesia." *Estuarine, Coastal and Shelf Science* 76 (1). Elsevier: 14–20.

Guest, JR, K Tun, J Low, A Vergés, EM Marzinelli, AH Campbell, AG Bauman, DA Feary, LM Chou, and PD Steinberg. 2016. "27 Years of Benthic and Coral Community Dynamics on Turbid, Highly Urbanised Reefs Off Singapore." *Scientific Reports* 6. Nature Publishing Group.

Hamilton, RJ, GR Almany, CJ Brown, J Pita, NA Peterson, and JH Choat. 2017. "Logging Degrades Nursery Habitat for an Iconic Coral Reef Fish." *Biological Conservation* 210. Elsevier: 273–80.

Hill, J, and CL Wilkinson. 2004. "Methods for Ecological Monitoring of Coral Reefs." Townsville: Australian Institute of Marine Science.

Hui, Francis KC. 2016. "Boral–Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in R." *Methods in Ecology and Evolution*. Wiley Online Library.

Hui, Francis KC, Sara Taskinen, Shirley Pledger, Scott D Foster, and David I Warton. 2015. "Model-Based Approaches to Unconstrained Ordination." *Methods in Ecology and Evolution* 6 (4). Wiley Online Library: 399–411.

Joseph, Maxwell B, Daniel L Preston, and Pieter TJ Johnson. 2016. "Integrating Occupancy Models and Structural Equation Models to Understand Species Occurrence." *Ecology* 97 (3). Wiley Online Library: 765–75.

Kleiber, Christian, and Achim Zeileis. 2016. "Visualizing Count Data Regressions Using Rootograms." *The American Statistician* 70 (3). Taylor & Francis: 296–303.

Kruskal, Joseph B. 1964. "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika* 29 (2). Springer: 115–29.

Legendre, Pierre, and Loic FJ Legendre. 2012. *Numerical Ecology*. Vol. 24. Elsevier.

Martin, Tara G, Brendan A Wintle, Jonathan R Rhodes, Petra M Kuhnert, Scott A Field, Samantha J Low-Choy, Andrew J Tyre, and Hugh P Possingham. 2005. "Zero Tolerance Ecology: Improving Ecological Inference by Modelling the Source of Zero Observations." *Ecology Letters* 8 (11). Wiley Online Library: 1235–46.

Plummer, Martyn. 2016. *Rjags: Bayesian Graphical Models Using Mcmc*. https://CRAN.R-project.org/package=rjags.

Plummer, Martyn, and others. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 124:125. Vienna.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Socolar, Jacob B, James J Gilroy, William E Kunin, and David P Edwards. 2016. "How Should Beta-Diversity Inform Biodiversity Conservation?" *Trends in Ecology & Evolution* 31 (1). Elsevier: 67–80.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2016. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and Waic." *Statistics and Computing*, 1–20.

Warton, David I, F Guillaume Blanchet, Robert B O'Hara, Otso Ovaskainen, Sara Taskinen, Steven C Walker,

and Francis KC Hui. 2015. "So Many Variables: Joint Modeling in Community Ecology." *Trends in Ecology & Evolution* 30 (12). Elsevier: 766–79.

Warwick, Richard Martyn. 1993. "Environmental Impact Studies on Marine Communities: Pragmatical Considerations." *Austral Ecology* 18 (1). Wiley Online Library: 63–80.

Warwick, Richard Martyn, and KR Clarke. 1991. "A Comparison of Some Methods for Analysing Changes in Benthic Community Structure." *Journal of the Marine Biological Association of the United Kingdom* 71 (01). Cambridge Univ Press: 225–44.