

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

Journal:	<i>Conservation Biology</i>
Manuscript ID	17-382.R2
Wiley - Manuscript type:	Contributed Paper
Keywords:	Ecological indicator, Coral reef, water quality, latent variable, Bayesian modelling, logging, multi-species distribution modelling, beta-diversity
Abstract:	<p>Ecological communities typically change along gradients of human impact, though it is difficult to estimate the footprint 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. The model estimates both changes in abundances of benthic groups and their compositional turn-over, a type of beta-diversity. 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 footprint of human impacts on ecosystems and evaluate the benefits of conservation actions for ecosystems.</p>

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

Abstract

Ecological communities typically change along gradients of human impact, though it is difficult to estimate the footprint 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. The model estimates both changes in abundances of benthic groups and their compositional turn-over, a type of beta-diversity. 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 footprint of human impacts on ecosystems and evaluate the benefits of conservation actions for ecosystems.

Introduction

Determining the extent and impact of diffuse threats, like pollution, 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 & Clarke 1991; De'ath & Fabricius 2010; DeMartini et al. 2013). 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).

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 abundance of organisms that make-up a community, including both species and higher taxonomic groupings. We will also refer to measurements made on counts of individuals, occurrences of a taxonomic group or per cent cover collectively as 'abundances'. The analysis of community change over physical gradients presents many technical challenges. In particular, there are often many species sampled across relatively few sites. Algorithms for the analysis of multivariate abundances help overcome this problem (e.g. Kruskal 1964; Anderson et al. 2011; Legendre & 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 significant differences in community composition, they cannot predict abundances at unsampled sites (Warton et al. 2015). Prediction at unsampled sites is crucial for estimating the extent of a diffuse threat when measurements are patchy.

Multivariate statistical models of abundances can directly model the effects of environmental covariates and thus enable prediction of community composition at unsampled sites (Warton et al. 2015). These new methods, termed 'joint modeling' build on generalized 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). Joint models can also be used to estimate

uncertainty in species-environment relationships. Abundance models may commonly have increased power to detect change in abundances when compared to traditional ordination methods, making them useful for detecting the response of rare species to human disturbances. However, joint models are a relatively recent development (Warton et al. 2015), so their application to conservation issues has been limited to date.

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 generalization 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 et al. 2016), or as a Bayesian analog to constrained factor analysis. The constrained latent variable reflects an unobserved gradient of community change at different distances from sediment sources. 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 modeling in conservation, we apply the model to survey data of coral reef communities in the Kia region of Solomon Islands. Sediment run-off from logging operations around Kia 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 aim to estimate how much more reef has been affected by these illegal operations. More generally, our objective is to use the case-study to illustrate how joint modeling can be applied to estimate the impact of new developments, or the benefits of conservation actions.

Methods

First we describe a general framework for modeling 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.

Model framework

We modeled abundances of multiple response groups as a function of multiple latent variables (Figure 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 modeled as a function of a linear predictor:

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

where $\eta_{i,j}$ was the value of the linear predictor for group j at site i , θ_i was an offset term that accounted for variation in sampling intensity across sites, α_j is a group level intercept term. 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. Equation 1 is flexible in that it can be used with a range of

statistical distributions to model different types of observations including counts, occurrences or per cent cover data.

The latent variables were sampled from normal distributions:

$$\gamma_{i,k} \sim N(\kappa_{k,\cdot} x_{\cdot,i}, 1), \quad (2)$$

where there are K latent variables across the i sites, $\kappa_{k,\cdot}$ is a row vector of coefficients and $x_{\cdot,i}$ is a column vector of covariate values for each site. Latent variables for residual correlations have $\kappa_{k,\cdot} = 0$.

An alternative model would be to specify direct effects of the covariates on each group's abundance (Hui et al. 2015). 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 on 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 the environmental gradient. 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 composition 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 for the extent of turbid

waters. Taking the latent variable as a measure of turbidity also means we should interpret its coefficients ($\kappa_{k,j}$) as representing attenuation 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).

To aid interpretation we used the inverse-probit transformation to transform the constrained latent variable to a probability scale. On the probability scale, values near 0 or 1 represent communities that are at extreme ends of the composition gradient. Higher estimated values of the coefficient κ_k for an environmental gradient X mean the probability will approach 0 and 1 at either end of the gradient. If κ_k is near zero then community composition does not vary across the gradient and all values on the probability scale will be near 0.5.

Priors for species loadings ($\beta_{j,k}$) and covariate effects ($\kappa_{k,j}$) were specified as normal distributions. We use uninformative priors for covariate effects (mean = 0, variance = 1000) and vaguely informative priors for species loadings (mean = 0, variance = 20) to aid convergence (Hui 2016). Vaguely informative priors are appropriate here because we standardized the covariate data, so we do not expect variances of effects to be $\gg 1$.

Case-study

We model change in cover of benthic habitats across a gradient of sedimentation in the Kia District of Isabel Province, Solomon Islands (Figure 2). Kia district has a human population of about 2000 people (Solomon Island Government 2009), the historical extent of forests in our study region was approximately 79000 ha (Hansen et al. 2009), but logging in Kia district has removed 50% of primary forests over the last 15 years (Fig. 2). Typical operations haul logs to

log ponds, which are locations on the coast where mangroves are bulldozed so logs can be stored and then transported elsewhere. Log ponds result in significant sedimentation entering near-shore 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 the survey data were collected in 2013.

Previously the region-wide loss of nursery habitat was estimated 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 (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 (Appendix S1), across a range of distances from log ponds of 0.39km to 32 km (Hamilton et al. 2017). Sites were surveyed with a modified version of the point intercept method (Hill & 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 (Appendix S2). Benthic habitats were categorized into 31 categories as per English et al. (1994); however, we added categories for dead branching *Acropora* and dead branching

coral (Appendix S3). Water clarity was measured with a Secchi disc in deep lagoonal water adjacent to the survey site prior to commencing benthic surveys (Hamilton et al. 2017).

For analysis we aggregated the 31 categories into 17 focal groups (Appendix S3). Aggregation sped computations and emphasized change in the habitat types with which we were most interested. We also aggregated categories to avoid having numerous infrequently recorded categories that had many zeros and contain little information about environmental gradients..

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

$$p_{i,j} = \frac{e^{\pi_{i,j}}}{\sum_{j=1}^J e^{\pi_{i,j}}}, \quad (3)$$

where $\pi_{i,j}$ is the expectation for the count of habitat j at site i .

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

The first latent variable was modeled as a function of distance to the nearest log pond and flow strength. We standardized the distance metric by subtracting its mean divided by its standard deviation. Flow was classified as either mild or strong, 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% C.I. 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 in the model to account for residual correlations among habitat categories. We decided to use two latent variables because results from two latent variables can be easily visualized (Hui 2016). In additional simulations we refit the models with up to 4 unconstrained latent variables. We also fit a model with two unconstrained latent variables and no constrained latent variables and compared its fit to the model with a constrained latent variable. We used the widely applicable information criterion (WAIC) (Vehtari et al. 2016) and leave-one-out cross validation to compare model fits.

We checked model fits by comparing predicted to observed per cent covers for each habitat type and we also used rootograms to ensure counts were approximately Poisson distributed (Kleiber & Zeileis 2016). We calculated the bias and variance when comparing predicted to observed cover levels. Additionally, we used leave-one-out cross validation to evaluate the model's fit to each habitat type (Appendix S5). To confirm the Bayesian approach was consistent with existing methods, we also compare its ordination of habitats and sites to those obtained from two widely used approaches to multivariate analysis (Legendre & Legendre 2012). The first was non-metric multidimensional scaling which performs an unconstrained ordination of sites that we relate to the environmental covariates post-hoc. The second was constrained correspondence analysis, where the ordination of communities at sites was constrained by distance to the nearest log pond and flow conditions (methodological details are in Appendix S6).

We addressed our second objective of mapping the footprint of logging by predicting the constrained latent variable across the entire study region. We used over water distance to the nearest log pond as a covariate and assumed mild flow at all sites, which was the dominant condition at surveyed reefs. We then mapped the values of the latent variable onto a 0-1 probability scale. Finally, we estimated the area of reefs within probability bands of 95%, 75%, 50% and 25%, where higher values indicate community composition nearer to log-ponds. 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 recently occurred logging on the north-western Coast (three log ponds at: [7°26'11"S, 158°11'59"E], [7°25'42"S, 158°12'12"E] and [7°25'32"S, 158°12'4"E]) and then we calculated the area of reefs that are predicted to be affected by increased turbidity.

We used the JAGS program (Plummer & others 2003) for parameter estimation, run from the R programming environment (R Core Team 2016) with the package rjags (Plummer 2016). For each model we ran a single chain with 1 000 000 samples, thinning for every 40th sample. We used only a single chain to avoid issues with sign switching on the latent variables (Hui 2016). We checked the Hellinger distance statistic (Boone et al. 2014) to confirm model convergence. Data are provided in Appendices S1-3, JAGS model code is provided in Appendix S4. [Note to copy editor – we recommend inserting link to Github project that contains code and data here <https://github.com/cbrown5/BenthicLatent>].

Results

The model with a constrained latent variable and two unconstrained variables (Figure 3) had stronger support than the ordination that had no constrained latent variables (Table 1). Adding more unconstrained variables also increased the support for the model (Table 1), but simultaneously weakened the effect of the constrained latent variable (Figure S1). The cross-validation also supported models with 1-2 unconstrained latent variables for those habitats that were related to distance to log ponds. We therefore used the model that had two unconstrained and one constrained latent variable for further analysis.

A significant positive effect of minimum distance to log ponds on the constrained latent variable (median = 0.47, 0.12 to 0.81 95% C.I.s) was in the same direction as strong flow conditions (median = 1.91, 0.9 to 2.93 95% C.I.s). The median effect of strong flow was ~4 times greater than the median effect of minimum distance on the latent variable. Thus, we can interpret a change from mild to strong flow as having an equivalent effect on community structure as moving 2.46 km farther away from log-ponds.

We checked the nearest distance model for bias in its predictions of each habitat separately by comparing the observed versus expected values for habitat counts (Appendix S5). In general the model fitted most habitats with low bias. Bias was higher for habitats that had many zero observations, where the model tended to under-predict their counts in surveys. Results for the site ordinations and habitat ordinations from the Bayesian model were comparable to CCA and MDS, with the exception of Halimeda algae (Appendix S6). Halimeda algae was found to make a stronger contribution to the water quality gradient in the Bayesian model than the CCA (Figure S6). There was generally higher cover of Halimeda algae closer to log-ponds, but counts of this habitat were highly variable and included many zeros. The Poisson error structure used in the

Bayesian model may have been better able to model the noisy response of *Halimeda* algae to log ponds than the CCA.

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 soft sediment, *Halimeda* algae, dead corals, macro-algae, turf algae and massive corals closer to log-ponds (Figures 3a, 4). At sites far from log ponds, branching *Acropora* were commonly observed growing out of sand, so the high cover of soft sediments but low cover of *Acropora* at sites near to log ponds ostensibly indicates the loss of *Acropora*. Multiple algae groups were also associated with sites closer to log ponds (Figure 4a); 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%) so that sites nearer to log ponds were dominated by sand and massive corals (Fig. 4).

The unconstrained latent variables indicated positive associations among cover of macro-algae and *Acropora* (Figure 3b) and a gradient from algal assemblages, rock and turf algae to branching coral, dead coral and tabulate *Acropora* (Fig. 3c).

Means estimated for the constrained latent variable at each site were negatively correlated with Secchi depth ($\rho = -0.42$, $p < 0.01$), 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 the footprint of logging impacts on reefs as the probability a reef at a given distance will have a degraded

community state similar to reefs nearer to log-ponds. We estimated that 0% (95% C.I.s also = 0) of reefs had a 95% probability, 0% (0 - 36% C.I.) of reefs had a 75% probability and 60% (60 - 60% C.I.) of reefs had a 50% probability. The low rate of affected reefs at high probability intervals (i.e. 95%) indicated that there were some unaffected reefs even in areas nearer log-ponds. Overall, most lagoonal reefs on the southern side of have been affected by log ponds with >50% probability (Fig. 2), however the weaker effect for higher probability intervals indicates considerable variability in reef state that was not explained by distance to log-ponds.

We also estimated the area of lagoonal reef that has been protected by actions to stop logging on the north-western side of Kia. We predicted that the inclusion of new log ponds affected 65% (65 - 66 95% C.I.) of reefs with 50% probability, amounting to an extra 5 % of reefs or up to 179 hectares of reefs.

Discussion

We used joint modeling 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, algae and unconsolidated sediment. This gradient is consistent with studies of turbidity gradients from other regions of the Pacific. For instance massive corals are more tolerant of low light conditions and high sediment loads (Fabricius 2005) than branching and foliose corals. The reefs of Hawaii, 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; DeMartini et al. 2013; Brown et al. 2017a). We also observed that algae cover did not increase at degraded sites and replace coral cover, as it may do when there are alternate ecological states (Roff & Mumby 2012). Our findings

are consistent with studies of Pacific reefs (Guest et al. 2016) where overall algal cover is generally low (Roff & Mumby 2012).

The estimated rate of change in the constrained latent variable can be interpreted in two ways. First, it represents a measure of directional turn-over across an environmental gradient and can thus be interpreted 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. Comparing beta-diversity may be useful, for instance, because it is related to temporal stability of communities (Mellin et al. 2014). The estimated rate of change could also be interpreted as the rate of attenuation 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 (e.g. Brown et al. 2017b). Because communities integrate environmental change over time they may therefore offer a more rapid way to assess environmental change than measuring pollution directly (Warwick & Clarke 1991), which can be highly variable through time.

The joint modeling used here enables prediction of abundances at unmeasured sites with uncertainty bounds (Warton et al. 2015). We estimated community structure across all of Kia's lagoonal reefs and found that ~60% of the region's reef communities have been affected by existing log-ponds with 50% probability. We can be confident that most southern reefs are affected by log ponds because logging in the southern region has been extensive. 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). 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 (Graham & Nash 2013, Sale & Hixon 2014). Pollution, in particular, low water clarity from sediment run-off is a key contributor to the decline of structurally complex reefs (Rogers 1990). Our joint modeling approach thus provides an important tool for assessing the footprint of human activities that have increased turbidity and sedimentation on reefs.

Models that predict 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 occurred in an area with very little lagoonal reef nearby. Although turbidity on the fringing reefs directly adjacent to the illegal logging ponds was very high in 2015 and 2016 (pers. Obs). We have not extrapolated our model to the nearby fringing reefs, because we do not have survey data for this type of reef. The broad width of the uncertainty limits suggests that more field surveys are required to characterize the footprint size of individual log-ponds. Quantification of uncertainty can be used by managers to weigh up the value of further surveys, which can be costly and time-consuming, against the need for rapid management decisions (McDonald-Madden et al. 2010). For instance, time-series data may provide more accurate, but more expensive estimation of logging impacts than the 'space-for-time' substitution we used here (Pickett 1989, DeMartini et al. 2013). In the case of Kia, local communities decided to prevent further logging at the northern end of their sea estates, because the risks posed by additional logging were greater than the potential benefits. Therefore in Kia, there has been no need to obtain more precise estimates for the footprint of individual log-ponds. If communities wished to strategically allow logging in specific areas that were least likely to affect reefs, we would

recommend additional monitoring be undertaken so the footprint could be more accurately estimated.

The benefits of quantifying the footprint of human development, and uncertainty in its size, highlight several advantages of joint modeling of communities when compared to the traditional approach of distance-based multivariate statistics (e.g. Anderson et al. 2011). Joint models of abundance can be used to predict abundance across the pollution gradient, whereas distance-based methods cannot predict changes in abundances directly. Predictions of abundance can be more directly interpreted than changes in relative measures of community distance (Hui et al. 2015), making joint models useful for evaluating the expected impacts of new developments. The impacts of new developments on rare species are often of particular concern to conservation managers. Joint models have greater power than distance-based methods to detect change in rare species, because they can better distinguish between spatial variation in abundance and changes in mean abundance that are caused by environmental gradients (Hui et al. 2015). Joint models also enable the prediction of uncertainty bounds for abundances, community state and the footprint of human impacts. Characterizing uncertainty in conservation planning is important, because more precise estimates of impact imply that management can make more localized decisions, such as finer-scale spatial planning for where forests should be protected from logging.

There are some technical challenges to further development of joint models and their use for informing conservation decisions. Joint models currently require a higher degree of technical sophistication to implement than traditional distance based metrics, because the analyst needs to design the structure of the model a priori. For instance, we found that the inclusion of greater numbers of unconstrained variables hid the effect of turbidity gradient. Inclusion of too many

unconstrained variables may enhance the model's fit to the data, but come at the cost of lower generality – because unconstrained variables cannot be predicted to unsampled sites. Therefore, we recommend the analyst select the number of constrained and unconstrained variables a priori (Hui et al. 2015), but further work is needed to test how accurate models with multiple unconstrained variables are when predicting to new regions. We also found the Bayesian model was better able to attribute noisy gradients in habitat cover to pollution than traditional methods like CCA. Improved statistical power in the Bayesian models over traditional methods may be because they explicitly model the mean and variance of abundance counts (Hui et al. 2015). The availability of programming packages like BORAL (Hui 2016), which have similar usability to popular distance-based algorithms (e.g. Oksanen et al. 2017) will enhance the accessibility of joint modeling techniques and their utility for conservation planning.

When designing surveys and models to evaluate the impacts of development it is important to consider the possibility that measured covariates may be spatially confounded with unmeasured variables (Dormann 2007). For instance, if distance to log ponds was associated with a confounding environmental gradient 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 such important confounding gradients in this Kia study, since all sites surveyed were lagoonal fringing reefs that were sheltered from significant wave exposure. Future studies could extend our joint model to include multiple environmental gradients as multiple different constrained latent variables. If there are hypothesized but unmeasured gradients, models of spatial autocorrelation can help obtain more accurate inference. Preliminary simulation studies suggest that inferences made from joint models may be relatively unbiased by missing covariates

(Warton et al. 2015). However, future methodological research should examine the effects of unmeasured covariates on the accuracy of estimates of the impacts of development.

Our flexible joint modeling approach to estimating the impact of logging on lagoonal coral reef communities enabled prediction of community turnover across a gradient of human impacts and of the extent of logging impacts to coral reefs. More generally, joint models offer a useful tool for assessing the footprint of diffuse human impacts on ecosystems and for evaluating the impact of future development proposals. This will help conservation decision makers balance the opportunities for economic development with their impacts to ecosystems and make more informed decisions about development.

Supporting information

Site level data (Appendix S1), benthic survey data (Appendix S2), benthic variable names (Appendix S3), computer code (Appendix S4), model evaluation (Appendix S5) and comparison of the new method to existing methods (Appendix S6) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature cited

Anderson MJ et al. 2011. Navigating the multiple meanings of β diversity: A roadmap for the practicing ecologist. *Ecology letters* **14**:19–28.

Boone EL, Merrick JR, Krachey MJ. 2014. A Hellinger distance approach to MCMC diagnostics. *Journal of Statistical Computation and Simulation* **84**:833–849..

- Brown CJ, Jupiter SD, Lin H-Y, Albert S, Klein C, Maina JM, Tulloch VJD, Wenger AS, Mumby PJ. 2017a. Habitat change mediates the response of coral reef fish populations to terrestrial run-off. *Marine Ecology Progress Series* **576**:55-68.
- Brown CJ et al. 2017b. Tracing the influence of land-use change on water quality and coral reefs using a Bayesian model. *Scientific Reports* **7**:4740 DOI:10.1038/s41598-017-05031-7
- Bryan G. 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**:389–410.
- DeMartini E, Jokiel P, Beets J, Stender Y, Storlazzi C, Minton D, Conklin E. 2013. Terrigenous sediment impact on coral recruitment and growth affects the use of coral habitat by recruit parrotfishes (f. *Scaridae*). *Journal of coastal conservation* **17**:417–429.
- De'ath G, Fabricius K. 2010. Water quality as a regional driver of coral biodiversity and macroalgae on the great barrier reef. *Ecological Applications* **20**:840–850.
- Dormann CF. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* **16**:129–138.
- English S, Wilkinson C, Baker V. 1994. Survey manual for tropical marine resources. Australian Institute of Marine Science, Townsville.
- Fabricius KE. 2005. Effects of terrestrial runoff on the ecology of corals and coral reefs: Review and synthesis. *Marine Pollution Bulletin* **50**:125–146.
- Gelman A, Hill J. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge.

- Golbuu Y, Fabricius K, Victor S, Richmond RH. 2008. Gradients in coral reef communities exposed to muddy river discharge in Pohnpei, Micronesia. *Estuarine, Coastal and Shelf Science* **76**:14–20.
- Graham N, Nash K. 2013. The importance of structural complexity in coral reef ecosystems. *Coral Reefs* **32**:315–326.
- Guest J, Tun K, Low J, Vergés A, Marzinelli E, Campbell A, Bauman A, Feary D, Chou L, Steinberg P. 2016. 27 years of benthic and coral community dynamics on turbid, highly urbanised reefs off Singapore. *Scientific Reports* (**6**:20717) DOI: 10.1038/srep20717
- Hamilton R, Almany G, Brown C, Pita J, Peterson N, Choat J. 2017. Logging degrades nursery habitat for an iconic coral reef fish. *Biological Conservation* **210**:273–280.
- Hansen MC, et al. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**: 850–53. Data available on-line from: earthenginepartners.appspot.com/science-2013-global-forest. Accessed 27/09/2017
- Hill J, Wilkinson C. 2004. Methods for ecological monitoring of coral reefs. Australian Institute of Marine Science, Townsville.
- Hui FK. 2016. Boral–Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution* **7**:744–750.
- Hui FK, Taskinen S, Pledger S, Foster SD, Warton DI. 2015. Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution* **6**:399–411.
- Joseph MB, Preston DL, Johnson PT. 2016. Integrating occupancy models and structural equation models to understand species occurrence. *Ecology* **97**:765–775.

Kleiber C, Zeileis A. 2016. Visualizing count data regressions using rootograms. *The American Statistician* **70**:296–303.

Kruskal JB. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**:115–129.

Legendre P, Legendre LF. 2012. *Numerical ecology*. Elsevier, Amsterdam.

McDonald-Madden E, Baxter PW, Fuller RA, Martin TG, Game ET, Montambault J, Possingham HP. 2010. Monitoring does not always count. *Trends in Ecology & Evolution* **25**:547–550.

Mellin C, Bradshaw C, Fordham D, Caley M. 2014. Strong but opposing -diversity–stability relationships in coral reef fish communities. *Proceedings of the Royal Society of London B: Biological Sciences* **281**:20131993..

Oksanen J et al. 2017. *Vegan: Community ecology package*. Comprehensive R Archive Network. Accessed 11/2016. Available from <https://CRAN.R-project.org/package=vegan>.

Pickett, S. T. A. 1989. Space-for-time substitution as an alternative to long-term studies. In *Long-term Studies in Ecology* G.E. Likens, ed.), pp. 110–135. New York: Springer-Verlag

Plummer M. 2016. *rjags: Bayesian graphical models using MCMC*. Available from <https://CRAN.R-project.org/package=rjags>. Accessed 31/10/2016.

Plummer M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Page 125 in *Proceedings of the 3rd international workshop on distributed statistical computing*. Vienna.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <https://www.R-project.org/>. Accessed 31/10/2016.

Roff G, Mumby PJ. 2012. Global disparity in the resilience of coral reefs. *Trends in Ecology & Evolution* **27**:404–413.

Rogers CS. 1990. Responses of coral reefs and reef organisms to sedimentation. *Marine Ecology Progress Series*. **62**:185–202.

Sale PF, Hixon MA. 2014. Addressing the global decline in coral reefs and forthcoming impacts on fishery yields. Pages 7-18 in Bortone SA *Interrelationships Between Corals and Fisheries*. CRC Press, Boca Raton.

Socolar JB, Gilroy JJ, Kunin WE, Edwards DP. 2016. How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution* **31**:67–80.

Solomon Islands Government. 2009. Census 2009: Basic Tables and Census Description. Honiara, Solomon Islands. www.statistics.gov.sb/statistics/demographic-statistics/census. Accessed 27/09/2017.

Vehtari A, Gelman A, Gabry J. 2016. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**:1413–1432.

Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FK. 2015. So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution* **30**:766–779.

Warwick RM. 1993. Environmental impact studies on marine communities: Pragmatical considerations. *Austral Ecology* **18**:63–80.

Warwick RM, Clarke K. 1991. A comparison of some methods for analysing changes in benthic community structure. *Journal of the Marine Biological Association of the United Kingdom* **71**:225–244.

For review only

Tables

Table 1 Comparison of candidate models using the widely applicable information criterion (WAIC). The Bayesian ordination includes only unconstrained latent variables. LV: number of unconstrained latent variables.

Model	Conditional WAIC
Bayesian ordination, 2 LVs	15621
Constrained model, 1 LV	15679
Constrained model, 2 LVs	12123
Constrained model, 3 LVs	10093
Constrained model, 4 LVs	8243

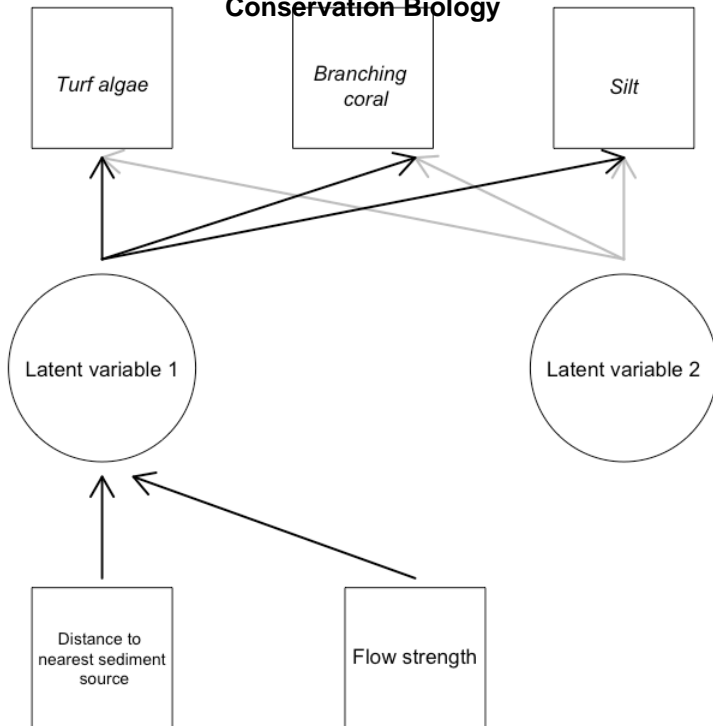
Figure Legends

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, circles indicate latent variables, variables in italics also have error terms that are estimated from the data. Arrows indicate model effects, with gray and black arrows indicating the effects relating to different latent variables.

Figure 2 Map of study region showing log ponds, survey sites and probabilities that a reef is degraded. The spatial field shows the probability that benthic communities are degraded. The probabilities are predicted to all ocean areas (not just lagoonal reefs) for visualization purposes. Note that uncertainty on the probability field is broad, but is not shown here for visualization purposes. The inset shows the location of the study region with reference to Honiara.

Figure 3 Mean estimates and credible intervals (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 farther from log-ponds.

Figure 4 Cover against distance to nearest log-ponds for habitats with a significant response to the latent variable for water quality. Significance was defined as 95% credibility intervals (C.I.s) not overlapping zero. Lines show median estimate and shaded areas show 95% C.I.s, points show observations.



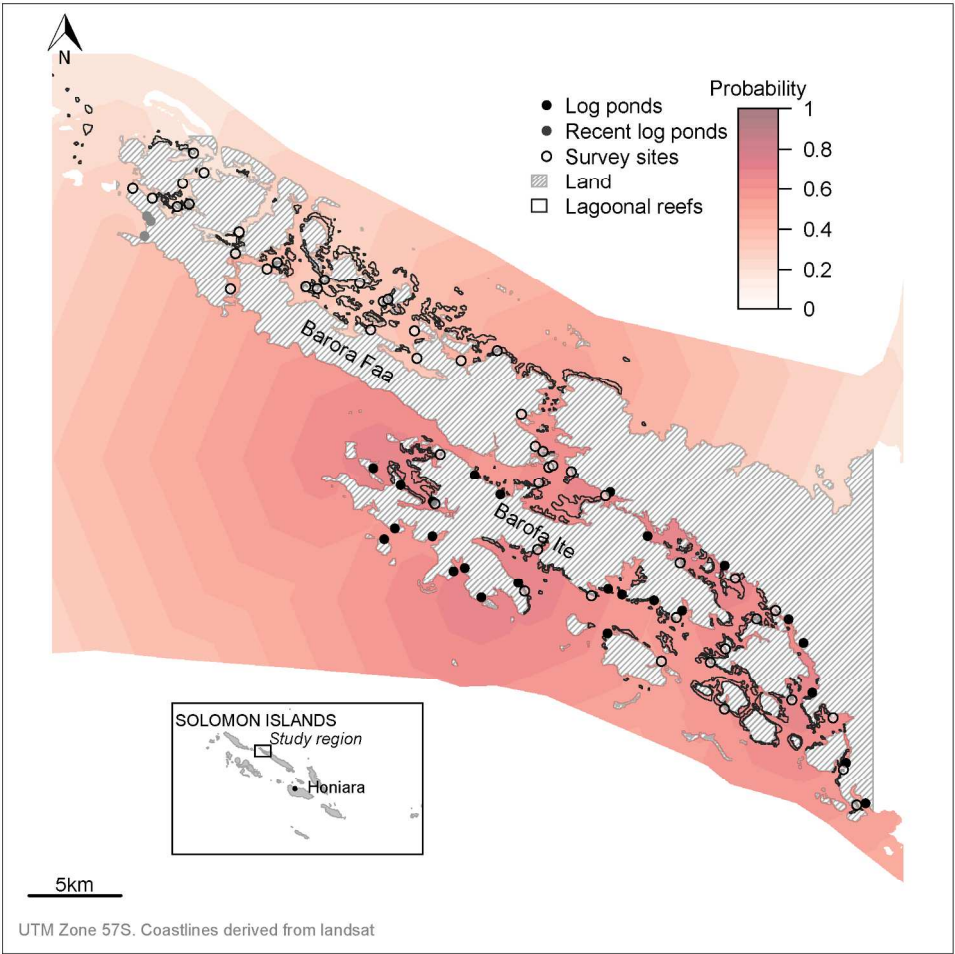


Figure 2 Map of study region showing log ponds, survey sites and probabilities that a reef is degraded. The spatial field shows the probability that benthic communities are degraded. The probabilities are predicted to all ocean areas (not just lagoonal reefs) for visualization purposes. Note that uncertainty on the probability field is broad, but is not shown here for visualization purposes. The inset shows the location of the study region with reference to Honiara.

846x846mm (72 x 72 DPI)

