

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

Journal:	<i>Conservation Biology</i>
Manuscript ID	17-382
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>

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

3    **Abstract**

4    Ecological communities typically change along gradients of human impact, though it is difficult to  
5    estimate the footprint of impacts for diffuse threats like pollution. Here we develop a joint model  
6    of benthic habitats on lagoonal coral reefs and use it to infer change in benthic composition along  
7    a gradient of distance from logging operations. The model estimates both changes in abundances  
8    of benthic groups and their compositional turn-over, a type of beta-diversity. We detect  
9    compositional turnover across the gradient and use the model to predict the footprint of  
10    turbidity impacts from logging. We then apply the model to predict impacts of recent logging  
11    activities, finding recent impacts to be small, because recent logging has occurred far from  
12    lagoonal reefs. Our model can be used more generally to estimate the footprint of human impacts  
13    on ecosystems and evaluate the benefits of conservation actions for ecosystems.

14    **Introduction**

15    Determining the extent and impact of diffuse threats, like pollution, is an important concern for  
16    ecological science and management. For instance, the extent of pollution impacts on marine  
17    benthic communities is often used to estimate the benefits of management actions aimed at its  
18    mitigation (Bryan 1971; Warwick & Clarke 1991; De'ath & Fabricius 2010; DeMartini et al. 2013).  
19    Evaluating the ecological benefits of mitigative efforts requires spatially comprehensive mapping  
20    of the ecosystems that are significantly affected. When the impact of a threat on different species  
21    varies, changes in community composition can be used to infer the spatial extent of a diffuse

22 threat (Warwick 1993). For instance, coastal development can increase turbidity of waters,  
23 causing shifts in coral reef communities toward stress tolerant species (e.g. Guest et al. 2016).  
24 Thus, species turnover, or beta-diversity, can be used to determine how far along a gradient one  
25 must move before the community is significantly changed (Anderson et al. 2011; Socolar et al.  
26 2016).

27 The spatial extent of diffuse impacts to communities can be challenging to infer from patchy  
28 measurements. Here we refer to 'community composition' as any multivariate measure of the  
29 abundance of organisms that make-up a community, including both species and higher  
30 taxonomic groupings. We will also refer to measurements made on counts of individuals,  
31 occurrences of a taxonomic group or per cent cover collectively as 'abundances'. The analysis of  
32 community change over physical gradients presents many technical challenges. In particular,  
33 there are often many species sampled across relatively few sites. Algorithms for the analysis of  
34 multivariate abundances help overcome this problem (e.g. Kruskal 1964; Anderson et al. 2011;  
35 Legendre & Legendre 2012). However, current methods are limited in that they are typically not  
36 based on models of abundances so, while they can be used to test for significant differences in  
37 community composition, they cannot predict abundances to unsampled sites (Warton et al.  
38 2015). Prediction at unsampled sites is crucial for estimating the extent of a diffuse threat when  
39 measurements are patchy.

40 Multivariate statistical models of abundances can directly model the effects of environmental  
41 covariates and thus enable prediction of community composition at unsampled sites (Warton et  
42 al. 2015). These new methods, termed 'joint modelling' build on generalised linear models to  
43 allow inclusion of multiple species and their interactions with each other and environmental  
44 covariates (Hui et al. 2015; Warton et al. 2015). Joint models can also be used to estimate

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

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

62 To illustrate the utility of joint modelling in conservation, we apply the model to survey data of  
63 coral reef communities in the Kia region of Solomon Islands. Around Kia sediment run-off from  
64 logging operations has degraded lagoonal reefs (Hamilton et al. 2017). We applied the joint  
65 model to surveys of lagoonal reef communities that were conducted across a gradient of  
66 distances from logging operations. We aimed to address three objectives in the case-study. First,  
67 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 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.

### 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  $\theta_i$  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  $i$ ,  $\alpha_j$  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. 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:

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

Where there are  $K$  latent variables across the  $i$  sites,  $\kappa_k$  is a vector of coefficients and  $X$  is a matrix of covariate values across the sites. Latent variables for residual correlations have  $\kappa_k = 0$ .

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 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 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 ( $\kappa_k = 0$ ) 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 transform the constrained latent variable to a 0-1 scale using the probit function. On the probit scale, values near 0 or 1 represent communities that are at extreme ends of the composition gradient. Higher estimated values of the coefficient  $\kappa_k$  for an environmental gradient  $X$  mean the probit transform will approach 0 and 1 at either end of the gradient. If  $\kappa_k$  is near zero then community composition does not vary across the gradient and all values on the probit scale will be near 0.5.

Priors for species loadings ( $\beta_{j,k}$ ) and covariate effects ( $\kappa_k = 0$ ) are specified using 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 standardised data, so we do not expect variances or 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. Logging in Kia district has removed 50% of forests over the last 15 years. Typical operations haul logs to log ponds, which are locations on the coast

132 where mangroves are bulldozed so logs can be stored then transported elsewhere. Log ponds  
133 result in significant sedimentation entering nearshore environments, which smothers corals and  
134 causes declines in branching corals and fish that depend on those corals for habitat (Hamilton et  
135 al. 2017). The southern part of the district has extensive logging (Hamilton et al. 2017), whereas  
136 the northern section was unlogged when the survey data were collected in 2013.

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

146 The data include surveys of benthic communities at 49 sites, conducted in October 2013  
147 (Hamilton et al. 2017). Sites were surveyed using a modified version of the point intercept  
148 method (Hill & Wilkinson 2004). Benthic habitats were recorded every 2 m along a 50 m  
149 transect, taking three measurements at each point, which were located directly below and one  
150 metre to the left and right of the transect. There were five transects of 75 points at each site,  
151 making a total of 375 points per site. Benthic habitats were categorised as per English et al.  
152 (1994), however we added categories for *Acropora* Branching Dead and Coral Branching Dead.  
153 Water clarity was also measured in deep water near to survey sites using a Secchi disc.



154 For analysis we aggregated categories into 17 focal groups (Appendix S1). Aggregation sped  
 155 computations and emphasised change in the habitat types we were most interested in. We also  
 156 aggregated categories to avoid having numerous infrequently recorded categories that had many  
 157 zeros. Rare categories would require additional modifications of the model to allow for over-  
 158 dispersion (e.g. zero-inflated models Martin et al. (2005)), which is beyond the scope of this  
 159 initial analysis.

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

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

167 *Equation 3*

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

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

173 The first latent variable was modelled as a function of distance to the nearest log pond and flow  
174 strength. We standardised the distance metric by subtracting its mean dividing by its standard  
175 deviation. Flow was classified as either low or high, based on experience of the survey divers.  
176 Previous analysis indicated that flow strength was important for explaining cover of branching  
177 corals (Hamilton et al. 2017). For all covariates we estimated mean effects with 95% credibility  
178 intervals, where a 95% CI that does not overlap zero is statistically significant. As an additional  
179 test of the model we compared the estimates of the sedimentation latent variable to Secchi  
180 depths that were measured during surveys (Hamilton et al. 2017).

181 We also included two unconstrained latent variables in the model to account for residual  
182 correlations among habitat categories. We decided to use two latent variables, because results  
183 from two latent variables can be easily visualised (Hui 2016). In additional simulations we refit  
184 the models with up to 4 unconstrained latent variables. We also fit a model with two  
185 unconstrained latent variables and no constrained latent variables and compared its fit to the  
186 model with a constrained latent variable using the Watanabe-Akaike information criterion  
187 (WAIC) (Vehtari et al. 2016)

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

191 We addressed our second objective of mapping the footprint of logging by predicting the  
192 constrained latent variable across the entire study region, using as a covariate the over water  
193 distance to the nearest log pond. Because we did not have flow estimates for the entire study  
194 region, we assumed low flow at all sites, which was the dominant condition at surveyed reefs. We  
195 then applied the probit link function to map the values of the latent variable onto a 0-1

196 probability scale. Finally, we estimated the area of reefs within probability bands of 95%, 75%,  
197 50% and 25%, where higher values indicate community composition closer to log-ponds. For  
198 reefs that were outside the range of distances from log ponds observed in the surveys we used  
199 the minimum and maximum observed distances to predict the condition variable, which avoided  
200 extrapolation.

201 Our final objective was to estimate the area of lagoonal reefs affected by new logging operations.  
202 We repeated the calculations for the area of reef affected by logging with new logging that has  
203 recently occurred logging on the north-western Coast (three log ponds at: [7.43639<sup>oS</sup>,  
204 158.19973<sup>oE</sup>], [7.42830<sup>oS</sup>, 158.20321<sup>oE</sup>] and [7.42558<sup>oS</sup>, 158.20111<sup>oE</sup>]) and then we calculated  
205 the area of reefs that are predicted to be affected by increased turbidity.

206 We used JAGS (Plummer & others 2003) for parameter estimation, run from the R programming  
207 environment (R Core Team 2016) using the package rjags (Plummer 2016). For each model we  
208 ran a single chain with 1 000 000 samples, thinning for every 40<sup>th</sup> sample. We used only a single  
209 chain to avoid issues with parameter switching on the latent variables (Hui 2016). We checked  
210 the Hellinger distance statistic (Boone et al. 2014) to confirm model convergence. Data are  
211 provided in Appendices S1-3, JAGS model code is provided in Appendix S4.

## 212 Results

213 The model with a constrained latent variable and two unconstrained variables had stronger  
214 support than the ordination that had no constrained latent variables (Table 1). Adding more  
215 unconstrained variables also increased the support for the model (Table 1), but simultaneously  
216 weakened the effect of the constrained latent variable (Appendix S5). However, we proceed with  
217 analysis using the model that had two unconstrained and one constrained latent variables.

218 There was a significant positive effect of minimum distance to log ponds on the constrained  
219 latent variable (median = 0.47, 0.12 to 0.81 95% C.I.s), that was in the same direction as high flow  
220 conditions (median = 1.91, 0.9 to 2.93 95% C.I.s). The median effect of high flow was ~4 times  
221 greater than the median effect of minimum distance on the latent variable. Thus, we can interpret  
222 a change from low to high flow as having an equivalent effect on community structure as moving  
223 2.46 km further away from log-ponds.

224 We checked the nearest distance model for bias in its predictions of each habitat separately  
225 (Appendix S5). In general the model fitted most habitats with low bias. Bias was higher for  
226 habitats that had many zero observations, where the model tended to underpredict their counts  
227 in surveys.

228 The loadings of the habitats on the constrained latent variable indicated that it represents a  
229 gradient from high cover of branching *Acropora*, branching corals, dead branching corals, other  
230 habitats and rare corals far from log-ponds, towards higher cover of soft sediment, *Halimeda*  
231 algae, dead corals, macro-algae, turf algae and massive corals closer to log-ponds (figs 2a, 4). At  
232 sites far from log ponds, branching *Acropora* were commonly observed growing out of sand, so  
233 the high cover of soft sediments but low cover of *Acropora* at sites near to log ponds ostensibly  
234 indicates the loss of *Acropora*. Multiple algae groups were also associated with sites closer to log  
235 ponds (Fig 3a), however their overall change increase closer to log ponds was small (Fig 3),  
236 suggesting they maintained their cover across the gradient, rather than replacing other habitats.  
237 Overall, cover of algae was low (<10%) so that sites near to log ponds were dominated by sand  
238 and massive corals (Fig 3).

239 The unconstrained latent variables indicated a positive association among cover macro-algae and  
240 *Acropora* (Fig 2, latent variable 1) and a gradient from algal assemblages, rock and turf algae to  
241 branching coral, dead coral and tabulate *Acropora* (Fig 2, latent variable 2).

242 Mean estimates for the constrained latent variable at taken at each site were negatively  
243 correlated with Secchi depth ( $\rho = 0.42$ ,  $p < 0.01$ ), where lower values of the latent variable  
244 indicate benthos associated with more turbid waters. The relatively weak correlation may be due  
245 to high temporal variability in Secchi depths that were measured only once at each site and the  
246 displacement of Secchi measurements from survey sites to deep channels.

247 By predicting the values of the condition variable across all reefs, we can estimate footprint of  
248 logging impacts on reefs, as the probability a reef at a given distance will have a degraded  
249 community state similar to reefs near to log-ponds. We estimated that 0% (95% C.I.s also = 0) of  
250 reefs had a 95% probability, 0% (0 - 36% C.I.) of reefs had a 75% probability and 60% (60 - 60%  
251 C.I.) of reefs had a 50% probability. The low rate of affected reefs at high probability intervals (i.e.  
252 95%) indicated that there were some unaffected reefs even in areas close to log-ponds. Overall,  
253 most lagoonal reefs on the southern side of have been affected by log ponds with >50%  
254 probability (Fig 4), however the weaker effect for higher probability intervals indicates  
255 considerable variability in reef state that was not explained by distance to log-ponds.

256 We also estimated the area of lagoonal reef that has been protected by actions to stop logging on  
257 the north-western side of Kia. We predicted with the inclusion of new log ponds affected 65% (65  
258 - 65 95% C.I.) of reefs with 50% probability, amounting to an extra 5 % of reefs or 179 to  
259 hectares of reefs.

## 260 Discussion

261 We used joint modelling to estimate the ecological impacts of logging on coral reef communities.  
262 We found there was a gradient in reef communities with reefs closer to log ponds lacking  
263 complex coral habitats and instead having higher cover of massive corals, algae and  
264 unconsolidated sediment. This gradient is consistent with studies of turbidity gradients from  
265 other regions of the Pacific. For instance massive corals are more tolerant of low light conditions  
266 and high sediment loads (Fabricius 2005) than branching and foliose corals. The reefs of Hawaii,  
267 Palau and Fiji have also been observed to lack cover of structurally complex hard corals, like  
268 branching *Acropora*, closer to turbid rivers (Golbuu et al. 2008; DeMartini et al. 2013; Brown et  
269 al. in press). We also observed that algae cover did not increase at degraded sites and replace  
270 coral cover, as it may do when there are alternate ecological states (Roff & Mumby 2012).  
271 However, this observation is consistent with studies of Pacific reefs (Guest et al. 2016) where  
272 overall algal cover is generally low (Roff & Mumby 2012).

273 The estimated rate of change in the constrained latent variable can be interpreted in two ways.  
274 First, it represents a measure of directional turn-over across an environmental gradient and can  
275 thus be interpreted as a kind of beta-diversity (Anderson et al. 2011). If similar models were fit for  
276 different regions, or a single model for multiple gradients, then we could compare beta-diversity  
277 across those gradients or regions. Comparing beta-diversity may be useful, for instance, because  
278 it is related to temporal stability of communities (Mellin et al. 2014). The estimated rate of  
279 change could also be interpreted as the rate of attenuation in turbidity, if we take benthic  
280 communities as indicators of environmental change. It would be fruitful in future research to  
281 compare these models to direct measurements of turbidity, or numerical models of turbidity and  
282 compare the spatial dispersion of the threat (e.g. Brown et al. 2017). Because communities

283 integrate environmental change over time they may therefore offer a more rapid way to assess  
284 environmental change than measuring pollution directly (Warwick & Clarke 1991), which can be  
285 highly variable through time.

286 The joint modelling used here enables prediction of abundances at unmeasured sites with  
287 uncertainty bounds (Warton et al. 2015). We estimated community structure across all of Kia's  
288 lagoonal reefs and found that ~60% of the region's reef communities have been affected by  
289 existing log-ponds with 50% probability. We can be confident that most southern reefs are  
290 affected by log ponds because logging in the southern region has been extensive. Lagoonal reefs  
291 with structurally complex corals are nursery habitat for locally important fishery species, so  
292 community change across these reefs has likely affected recruitment of juveniles to the fishery  
293 and local livelihoods (Hamilton et al. 2017). Globally, there is concern that loss of structurally  
294 complex habitats will cause widespread declines in reef fish species and affect human  
295 communities that depend on reef fisheries (Graham & Nash 2013, Sale & Hixon 2014). Pollution,  
296 in particular low water clarity from sediment run-off is a key contributor to the decline of  
297 structurally complex reefs (Rogers 1990). Our joint modelling approach thus provides an  
298 important tool for assessing the footprint of human activities that have increased turbidity and  
299 sedimentation on reefs.

300 Models that predict abundance are useful for conservation planning because the benefits of  
301 conservation actions and the impacts of development on abundance can be quantified. We  
302 predicted that recent illegal logging on the north coast of Kia may have further degraded only a  
303 small area of lagoonal reefs (<5%). This finding was consistent with our intuition, because the  
304 new logging occurred in an area with very little lagoonal reef nearby. Although turbidity on the  
305 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). 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 modelling of communities when compared to the traditional approach of using distance-based multivariate statistics (e.g. Socolar et al. 2016). Joint models of abundance can be used to predict abundance across the pollution gradient, whereas distance-based methods cannot predict changes in abundance 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 change in mean abundance that are caused by environmental gradients (Hui et al. 2015). Joint models also enable the prediction of uncertainty bounds on abundances, community state and the footprint of human impacts. Characterizing uncertainty in conservation planning is



330 important, because more precise estimates of impact imply that management can make more  
331 precise decisions, such as precise spatial planning for where forests should be protected from  
332 logging.

333 There are some technical challenges to further development of joint models and their application  
334 to evaluating conservation success. Joint models currently require a higher degree of technical  
335 sophistication to implement than traditional distance based metrics, because the analyst needs to  
336 design the structure of the model a-priori. For instance, we found that the amount of variance in  
337 community abundances explained by the constrained latent variable decreased if we included  
338 more unconstrained latent variables. This occurred because the unconstrained variables have  
339 greater flexibility to explain the community abundances than the constrained variable. Therefore,  
340 we recommend the analyst select the number of constrained and unconstrained variables a-  
341 priori (Hui et al. 2015), but further work is needed to test the implications of this decision.  
342 However, in general the number of decisions that must be made when fitting a joint model is  
343 similar to those faced by an analyst using traditional distance-based measures, where decision  
344 must be made about the distance metric and its transformation (Anderson et al. 2011). The  
345 availability of programming packages like BORAL (Hui 2016), which have similar useability to  
346 popular distance-based algorithms (e.g. Oksanen et al. 2017) will enhance the accessibility of  
347 joint modelling techniques and their utility for conservation planning.

348 When designing surveys and models to evaluate the impacts of development it is important to  
349 consider the possibility that measured covariates may be spatially confounded with unmeasured  
350 variables (Dormann 2007). For instance, if distance to log ponds was associated with a  
351 confounding environmental gradient such as exposure to wind and waves, we could falsely  
352 attribute community change driven by exposure to turbidity. If this occurred, we would

353 overestimate both the impact of existing log-ponds and the impact of future logging. We do not  
354 know of any such important confounding gradients in this Kia study, since all sites surveyed were  
355 lagoonal fringing reefs that were sheltered from significant wave exposure. Where there are  
356 hypothesised but unmeasured spatial dependencies, models of spatial autocorrelation can help  
357 obtain more accurate inference. Preliminary simulation studies suggest that inferences made  
358 from joint models may be relatively unbiased by missing covariates (Warton et al. 2015).  
359 However, future methodological research should examine the effects of unmeasured covariates  
360 on the accuracy of joint models at estimating the impacts of development.

361 We estimated the impact of logging on coral reef communities. To do so, we developed a flexible  
362 joint modelling approach. This approach enabled prediction of community turnover across a  
363 gradient of human impacts and prediction of the extent of logging impacts to coral reefs. More  
364 generally, joint models offer a useful tool for assessing the footprint of diffuse human impacts on  
365 ecosystems and for evaluating the impact of future development proposals. This will help  
366 conservation decision makers balance the opportunities for economic development with their  
367 impacts to ecosystems and make more informed decisions about development.

## 368 **Literature cited**

369 Anderson MJ et al. 2011. Navigating the multiple meanings of  $\beta$  diversity: A roadmap for the  
370 practicing ecologist. *Ecology letters* **14**:19–28. Wiley Online Library.

371 Boone EL, Merrick JR, Krachey MJ. 2014. A hellinger distance approach to mcmc diagnostics.  
372 *Journal of Statistical Computation and Simulation* **84**:833–849. Taylor & Francis.

373 Brown C, S J, H-Y L, S A, C J K, J M M, V T, A W, P J M. in press. Habitat change mediates the response  
374 of coral reef fish populations to terrestrial run-off. *Marine Ecology Progress Series*. MEPS.

- 375 Brown C et al. 2017. Tracing the influence of land-use change on water quality and coral reefs  
376 using a bayesian model. Scientific Reports. Nature.
- 377 Bryan G. 1971. The effects of heavy metals (other than mercury) on marine and estuarine  
378 organisms. Proceedings of the Royal Society of London B: Biological Sciences **177**:389–410. The  
379 Royal Society.
- 380 DeMartini E, Jokiel P, Beets J, Stender Y, Storlazzi C, Minton D, Conklin E. 2013. Terrigenous  
381 sediment impact on coral recruitment and growth affects the use of coral habitat by recruit  
382 parrotfishes (f. scaridae). Journal of coastal conservation:417–429. JSTOR.
- 383 De'ath G, Fabricius K. 2010. Water quality as a regional driver of coral biodiversity and  
384 macroalgae on the great barrier reef. Ecological Applications **20**:840–850. Wiley Online Library.
- 385 Dormann CF. 2007. Effects of incorporating spatial autocorrelation into the analysis of species  
386 distribution data. Global ecology and biogeography **16**:129–138. Wiley Online Library.
- 387 English S, Wilkinson C, Baker V. 1994. Survey manual for tropical marine resources. Australian  
388 Institute of Marine Science, Townsville.
- 389 Fabricius KE. 2005. Effects of terrestrial runoff on the ecology of corals and coral reefs: Review  
390 and synthesis. Marine pollution bulletin **50**:125–146. Elsevier.
- 391 Gelman A, Hill J. 2006. Data analysis using regression and multilevel/hierarchical models.  
392 Cambridge university press.
- 393 Golbuu Y, Fabricius K, Victor S, Richmond RH. 2008. Gradients in coral reef communities exposed  
394 to muddy river discharge in pohnpei, micronesia. Estuarine, Coastal and Shelf Science **76**:14–20.  
395 Elsevier.

- 396 Graham N, Nash K. 2013. The importance of structural complexity in coral reef ecosystems. *Coral*  
397 *Reefs* **32**:315–326. Springer.
- 398 Guest J, Tun K, Low J, Vergés A, Marzinelli E, Campbell A, Bauman A, Feary D, Chou L, Steinberg P.  
399 2016. 27 years of benthic and coral community dynamics on turbid, highly urbanised reefs off  
400 singapore. *Scientific Reports* **6**. Nature Publishing Group.
- 401 Hamilton R, Almany G, Brown C, Pita J, Peterson N, Choat J. 2017. Logging degrades nursery  
402 habitat for an iconic coral reef fish. *Biological Conservation* **210**:273–280. Elsevier.
- 403 Hill J, Wilkinson C. 2004. Methods for ecological monitoring of coral reefs. Page 117. Australian  
404 Institute of Marine Science, Townsville.
- 405 Hui FK. 2016. Boral–Bayesian ordination and regression analysis of multivariate abundance data  
406 in r. *Methods in Ecology and Evolution*. Wiley Online Library.
- 407 Hui FK, Taskinen S, Pledger S, Foster SD, Warton DI. 2015. Model-based approaches to  
408 unconstrained ordination. *Methods in Ecology and Evolution* **6**:399–411. Wiley Online Library.
- 409 Joseph MB, Preston DL, Johnson PT. 2016. Integrating occupancy models and structural equation  
410 models to understand species occurrence. *Ecology* **97**:765–775. Wiley Online Library.
- 411 Kleiber C, Zeileis A. 2016. Visualizing count data regressions using rootograms. *The American*  
412 *Statistician* **70**:296–303. Taylor & Francis.
- 413 Kruskal JB. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*  
414 **29**:115–129. Springer.
- 415 Legendre P, Legendre LF. 2012. *Numerical ecology*. Elsevier.

- 416 Martin TG, Wintle BA, Rhodes JR, Kuhnert PM, Field SA, Low-Choy SJ, Tyre AJ, Possingham HP.  
417 2005. Zero tolerance ecology: Improving ecological inference by modelling the source of zero  
418 observations. *Ecology letters* **8**:1235–1246. Wiley Online Library.
- 419 McDonald-Madden E, Baxter PW, Fuller RA, Martin TG, Game ET, Montambault J, Possingham HP.  
420 2010. Monitoring does not always count. *Trends in Ecology & Evolution* **25**:547–550. Elsevier.
- 421 Mellin C, Bradshaw C, Fordham D, Caley M. 2014. Strong but opposing -diversity–stability  
422 relationships in coral reef communities. *Proceedings of the Royal Society of London B: Biological*  
423 *Sciences* **281**:20131993. The Royal Society.
- 424 Oksanen J et al. 2017. Vegan: Community ecology package. Available from [https://CRAN.R-](https://CRAN.R-project.org/package=vegan)  
425 [project.org/package=vegan](https://CRAN.R-project.org/package=vegan).
- 426 Plummer M. 2016. Rjags: Bayesian graphical models using mcmc. Available from [https://CRAN.R-](https://CRAN.R-project.org/package=rjags)  
427 [project.org/package=rjags](https://CRAN.R-project.org/package=rjags).
- 428 Plummer M, others. 2003. JAGS: A program for analysis of bayesian graphical models using gibbs  
429 sampling. Page 125 in *Proceedings of the 3rd international workshop on distributed statistical*  
430 *computing*. Vienna.
- 431 R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for  
432 Statistical Computing, Vienna, Austria. Available from <https://www.R-project.org/>.
- 433 Roff G, Mumby PJ. 2012. Global disparity in the resilience of coral reefs. *Trends in Ecology &*  
434 *Evolution* **27**:404–413. Elsevier.
- 435 Rogers CS. 1990. Responses of coral reefs and reef organisms to sedimentation. *Marine ecology*  
436 *progress series*. Oldendorf **62**:185–202.

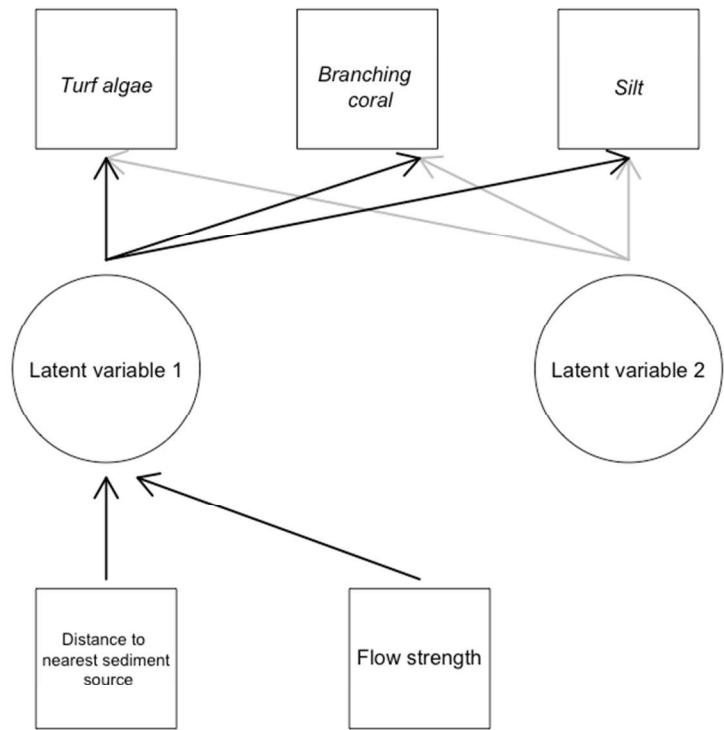
- 437 Sale PF, Hixon MA. 2014. Addressing the global decline in coral reefs and forthcoming impacts on  
438 fishery yields. *Interrelationships Between Corals and Fisheries* (ed. Bortone SA):7–18.
- 439 Socolar JB, Gilroy JJ, Kunin WE, Edwards DP. 2016. How should beta-diversity inform biodiversity  
440 conservation? *Trends in ecology & evolution* **31**:67–80. Elsevier.
- 441 Vehtari A, Gelman A, Gabry J. 2016. Practical bayesian model evaluation using leave-one-out  
442 cross-validation and waic. *Statistics and Computing*:1–20.
- 443 Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FK. 2015. So many  
444 variables: Joint modeling in community ecology. *Trends in ecology & evolution* **30**:766–779.  
445 Elsevier.
- 446 Warwick RM. 1993. Environmental impact studies on marine communities: Pragmatical  
447 considerations. *Austral Ecology* **18**:63–80. Wiley Online Library.
- 448 Warwick RM, Clarke K. 1991. A comparison of some methods for analysing changes in benthic  
449 community structure. *Journal of the Marine Biological Association of the United Kingdom*  
450 **71**:225–244. Cambridge Univ Press.

## Tables

*Table 1* Model comparison using the WAIC. LV: number of unconstrained latent variables.

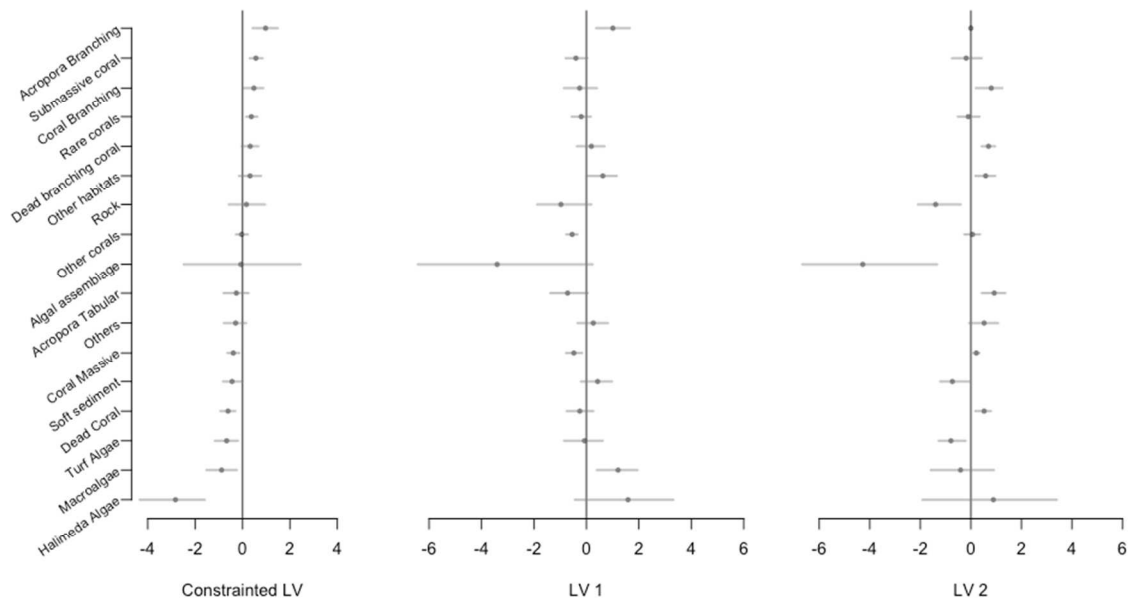
Model	WAIC
Bayesian ordination, 2 LVs	15621
Constrained model, 1 LV	15679
Constrained model, 2 LV	12123
Constrained model, 3 LV	10093
Constrained model, 4 LV	8243

Figures

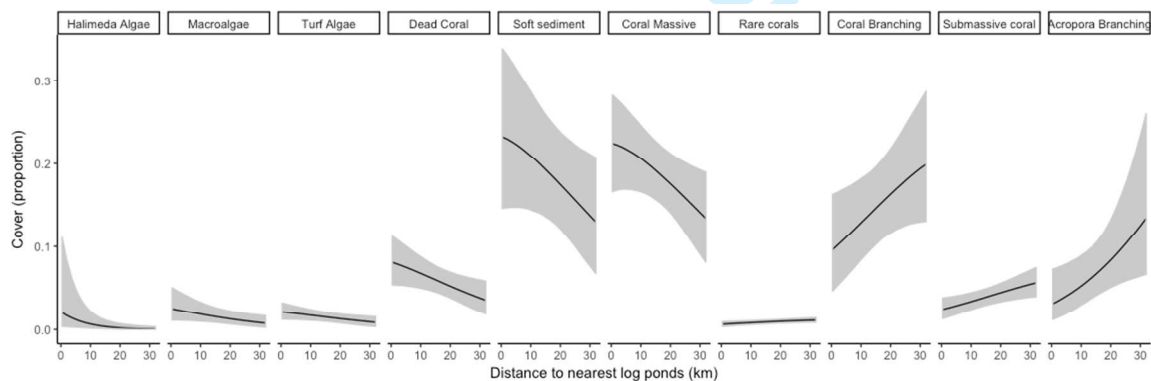


**Figure 1** Directed graph giving an example for the structure of our Bayesian latent variable model applied to coral reef habitats in Solomon Islands. Squares indicate measured variables, circles indicate latent variables, variables in italics also have error terms that are estimated from the data. Arrows indicate model effects, with gray and black arrows indicating the effects relating to different latent variables.

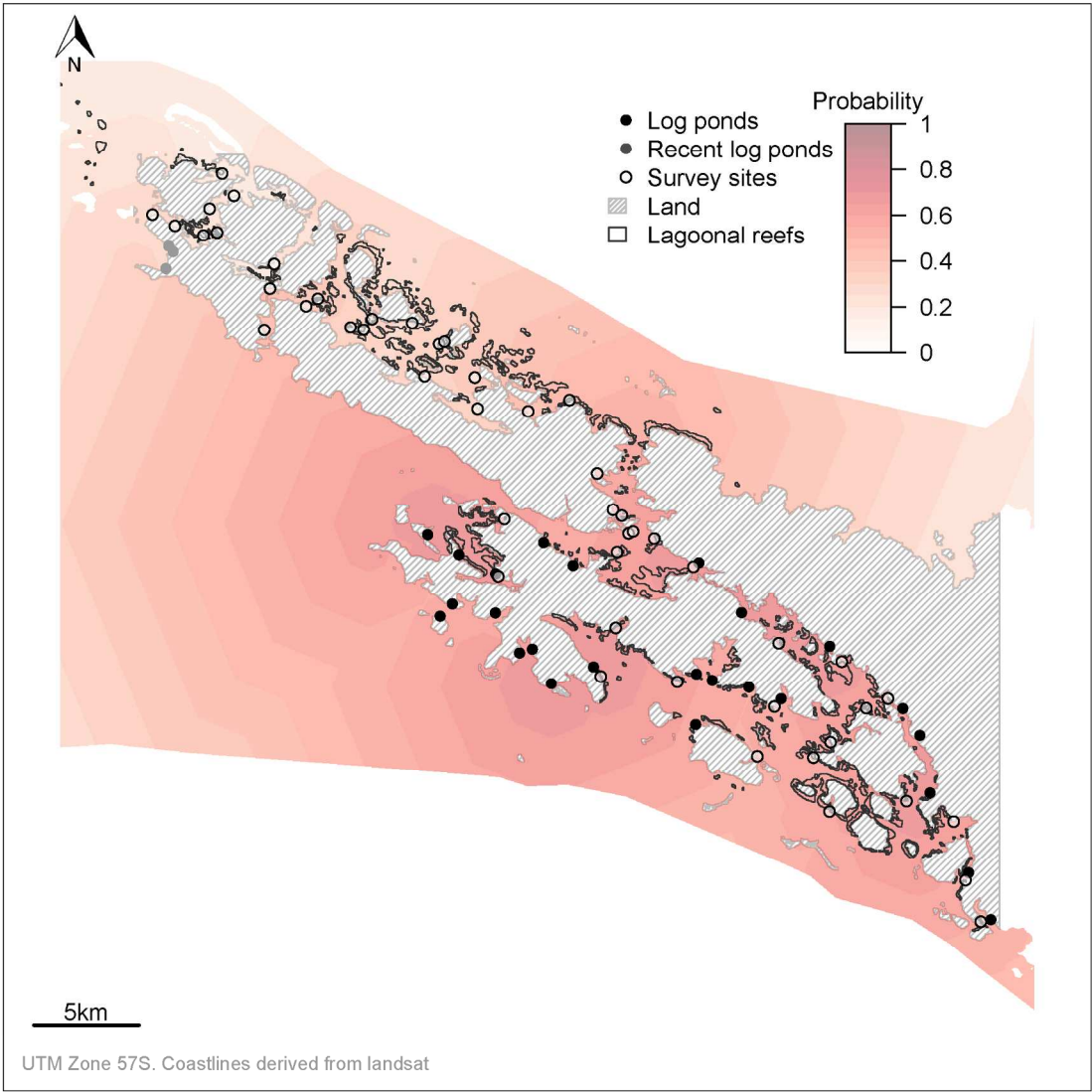




**Figure 2** Mean estimates and CIs for loadings of each habitat category on the (a) constrained latent variable and (b-c) two unconstrained latent variables. Loading signs on the constrained latent variable (a) are fixed so that positive values indicate a habitat's cover increases further from log-ponds.



**Figure 3** Cover against distance to nearest log-ponds for habitats with a significant response to the latent variable for water quality. Significant was defined as 95% CIs not overlapping zero.



**Figure 4** 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, calculated by taking a probit transform over the constrained latent variable. 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 is not shown here for visualisation purposes.