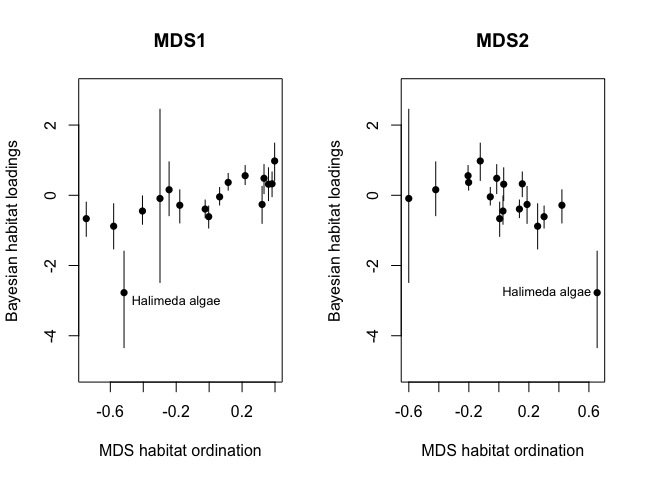
# Appendix S6 Comparison to existing multivariate analysis methods

Here we compare the results of our proposed latent variable model with those that would be obtained using some existing methods for analysis of multivariate community data and environmental gradients. We compare existing methods to the fits of the model with one constrained latent variable and two unconstrained latent variables.

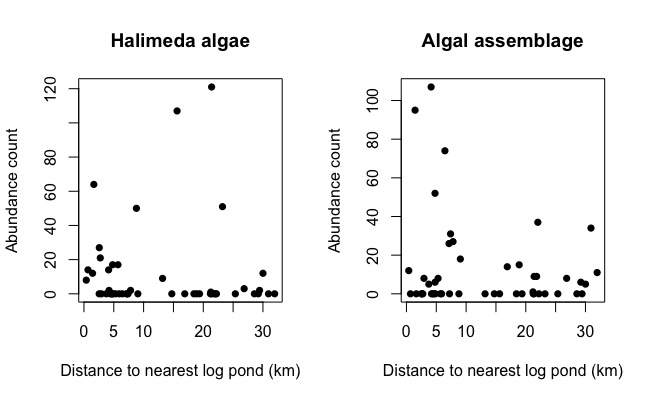
## Non-metric Multidimensional scaling

We conducted ordination on the benthic community data using a non-metric multidimensional scaling (MDS) with two dimensions, transforming the benthic abundance data using a sqrt transform to standardise the mean-variance relationship. We also tried a logit transform of proportional cover of each benthic group, however results were similar, so here we just present analysis of the abundance data. The MDS was performed using the vegan (Oksanen et al. 2017) package from the R programming language (R Core Team 2016). We then fit the environmental variables minimum distance to log ponds and flow (mild or strong) to the ordination using the envfit function, with 1000 permutations to evaluate the significance of each environmental predictor.

The MDS was able to resolve some patterns in community structure (stress parameter = 0.13). The fit of the MDS ordination to the environmental variables showed a significant (p = 0.003) relationship with flow conditions, but not minimum distance to log ponds (p = 0.71). Thus, the MDS was unable to detect the weaker effect of distance to log ponds on the community structure.  
Results of the ordination of habitats by the MDS were comparable to the Bayesian model, even though the MDS axes are unconstrained, suggesting that a gradient in community structure is strong enough to manifest without an a-priori hypothesis for its cause (Figure S2). A notable outlier was Halimeda algae, which was the Bayesian model classified as having a greater influence on the constrained latent variable relative to other habitats when compared to the MDS axes. Halimeda algae had higher cover closer to log-ponds, but also some very high cover values at sites far from log-ponds (Figure S3).



**Figure S2** Comparison of habitat ordination from the MDS with each the loadings of each habitat on the constrained latent variable from the Bayesian model. Note that the sign of each weight/loading is not comparable across the two methods. Points give median value for the Bayesian model and bars give 95% CIs.



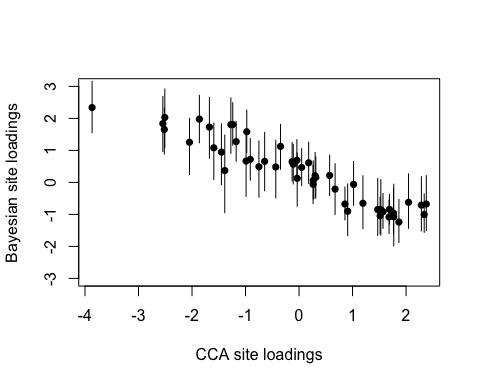
**Figure S3** Counts of the habitat types Halimeda algae and Algal assemblage at sites by their distance to the nearest log pond.

## Constrained correspondence analysis

We used constrained correspondence analysis (CCA) is to constrained an ordination of habitats structure across sites by environmental gradients. We conducted ordination on the benthic community data using constraints for minimum distance to log ponds and flow conditions. The CCA was computed using the vegan (Oksanen et al. 2017) package for the R programming language (R Core Team 2016)..

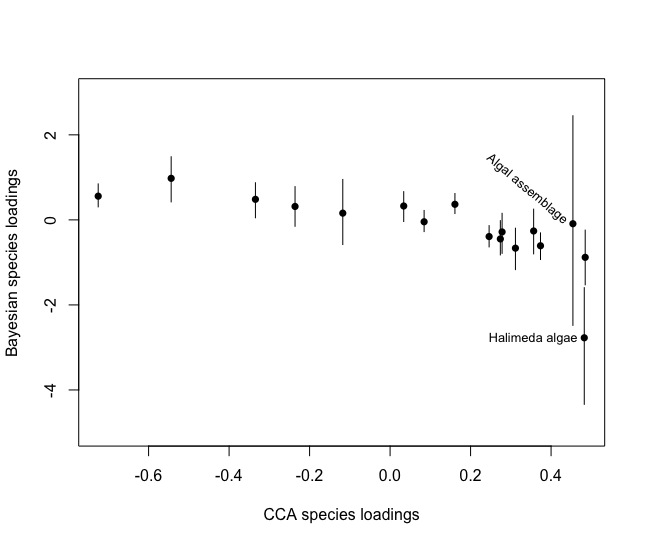
The two constrained axes accounted for 9.8% of the variation in community structure, with the first constrained axis accounting for 83% of that variance. Given the dominance of the first axis, we proceed with the comparison focussing on the first constrained axis from the CCA.

The CCA and the Bayesian ordination both identified a gradient of communities from sites close to log ponds that had low flow toward sites far from log ponds that had high flow. However, the signs of their respective ordinations were switched. For the CCA, the scores of the environmental predictors on the first axis were both negative (-0.19 for distance to log ponds and -0.85 for flow conditions), indicating that it represents a gradient of sites that had high flow or were far from log ponds at low negative values toward sites that had low flow or were closer to log ponds at high positive values. The ordination of sites on the first axis of the CCA and the ordination of sites from the constrained latent variable in the Bayesian model were strongly negatively related (Figure S4, r2 = 0.88). Higher positive values of the latent variable from the Bayesian model indicated sites that were further from log ponds and had higher flow, so the two ordinations represent the same gradient.

 **Figure S4** Comparison of site ordination from the CCA with the values for sites on the constrained latent variable from the Bayesian model. Note that the sign of each weight/loading is not comparable across the two methods. Points give median value for the Bayesian model and bars give 95% CIs.

The ordination of species from the CCA and the species loadings from the Bayesian model were also strongly related, indicating that both methods agree on which species were the dominant drivers of the gradient (Figure S5). However, once again the signs were switched and there was a negative relationship between the ordination of species across the environemtnal gradient for the two methods (r2 = 0.51).

A notable difference between the Bayesian model and the CCA was the role of algal assemblages (a habitat type) and Halimeda algae in determining the constrained latent variables. The Bayesian model indicated that the contribution of algal assemblages to the latent variable was near zero and highly uncertain (Fig S5), whereas on the CCA axis it had strong positive influence. Cover of algal assemblages was on average higher close to log-ponds, but there were also many sites with zero cover of this habitat close to log ponds (Fig S3). The CCA was unable to capture this variability. Halimeda algae was also higher at the sites close to log ponds, but its greatest values occurred far from log-ponds. The Bayesian model gave this habitat a very strong influence on the latent variable because the outlier sites with high values far from log-ponds were accounted for in other unconstrained latent variables.



**Figure S5** Comparison of species ordination from the CCA with the values for species on the constrained latent variable from the Bayesian model. Note that the sign of each weight/loading is not comparable across the two methods. Points give median value for the Bayesian model and bars give 95% CIs.

**Literature Cited**

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/>.

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, et al. 2017. *Vegan: Community Ecology Package*. <https://CRAN.R-project.org/package=vegan>.