# Appendix S6 Comparison to existing multivariate analysis methods

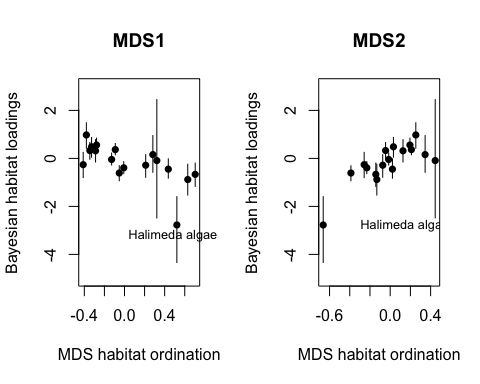
Here we compare the results of our proposed latent variable model with those that would be obtained using some existing methods for analysis of multivariate community data and environmental gradients. We compare existing methods to the fits of the model with one constrained latent variable and two unconstrained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

Run 0 stress 0.1288141   
Run 1 stress 0.1305837   
Run 2 stress 0.1317   
Run 3 stress 0.130584   
Run 4 stress 0.1288139   
... New best solution  
... Procrustes: rmse 0.0005542212 max resid 0.002124364   
... Similar to previous best  
Run 5 stress 0.1360768   
Run 6 stress 0.1308416   
Run 7 stress 0.1288163   
... Procrustes: rmse 0.001508446 max resid 0.006394262   
... Similar to previous best  
Run 8 stress 0.1417572   
Run 9 stress 0.1305875   
Run 10 stress 0.1307985   
Run 11 stress 0.1315816   
Run 12 stress 0.1288169   
... Procrustes: rmse 0.0005697746 max resid 0.002788336   
... Similar to previous best  
Run 13 stress 0.1360761   
Run 14 stress 0.1315829   
Run 15 stress 0.1305864   
Run 16 stress 0.1305853   
Run 17 stress 0.136077   
Run 18 stress 0.1353209   
Run 19 stress 0.1308447   
Run 20 stress 0.1360752   
\*\*\* Solution reached

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 from the MDS were comparable to the Bayesian model, even though the MDS is unconstrained, suggesting that a gradient in community structure is strong enough to manifest without an a-priori hypothesis for its cause. The species ordination for the two MDS axes were similar to loadings of species on the first constrained latent variable (Figure S2), even though the MDS was not constrained by any environmental predictors.



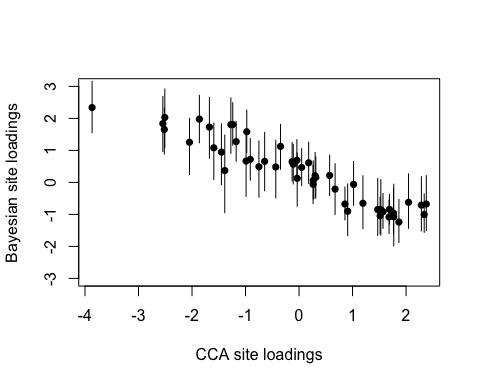
**Figure S2** Comparison of habitat ordination from the MDS with each the loadings of each habitat on the constrained latent variable from the Bayesian model. Note that the sign of each weight/loading is not comparable across the two methods. Points give median value for the Bayesian model and bars give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

## Constrained correspondence analysis

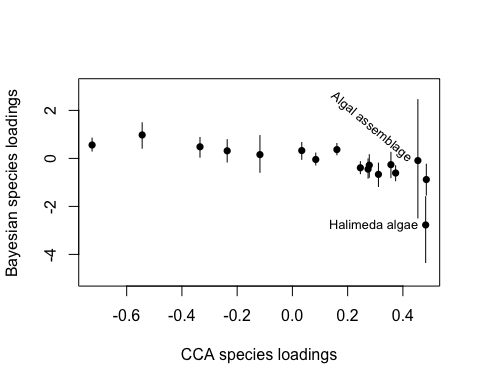
Constrained correspondence analysis (CCA) is used to constrained an ordination habitats across sites by environmental gradients. We conducted ordination on the benthic community data using constraints for minimum distance to log ponds and flow conditions. The CCA was computed using the vegan (Oksanen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

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.

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 S3, 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 S3** 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 S4). 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).

 **Figure S4** 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.

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>.