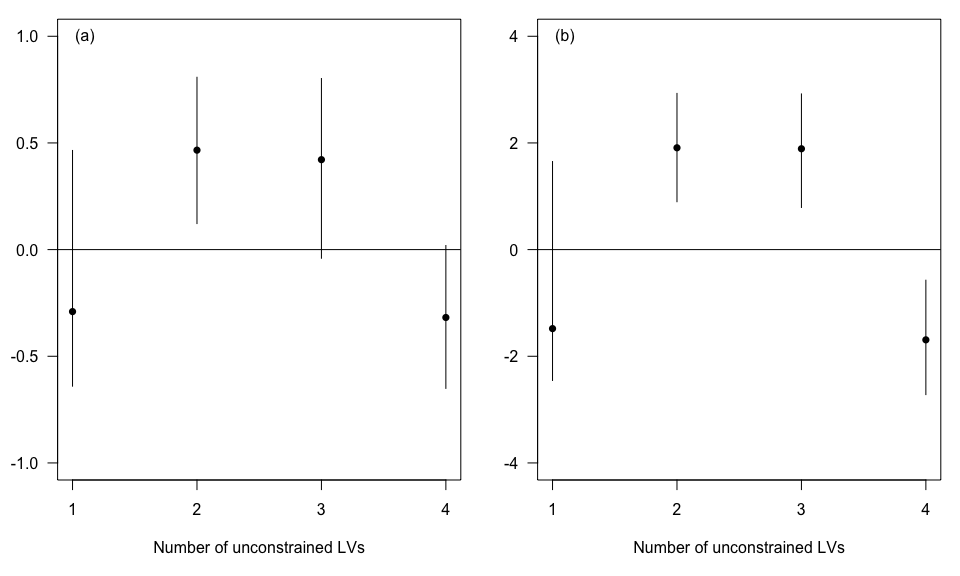
# Appendix S5 Model Evaluation

Here we conduct several additional analyses to confirm the model appropriately fits the data. We compared the estimates and credibility intervals for the effects of log-ponds and flow conditions across models with 1-4 unconstrained latent variables (Figure S1). The model was best able to detect an effect of log-ponds and flow conditions with 2 unconstrained latent variables, as indicted by 95% CIs that do not overlap zero.

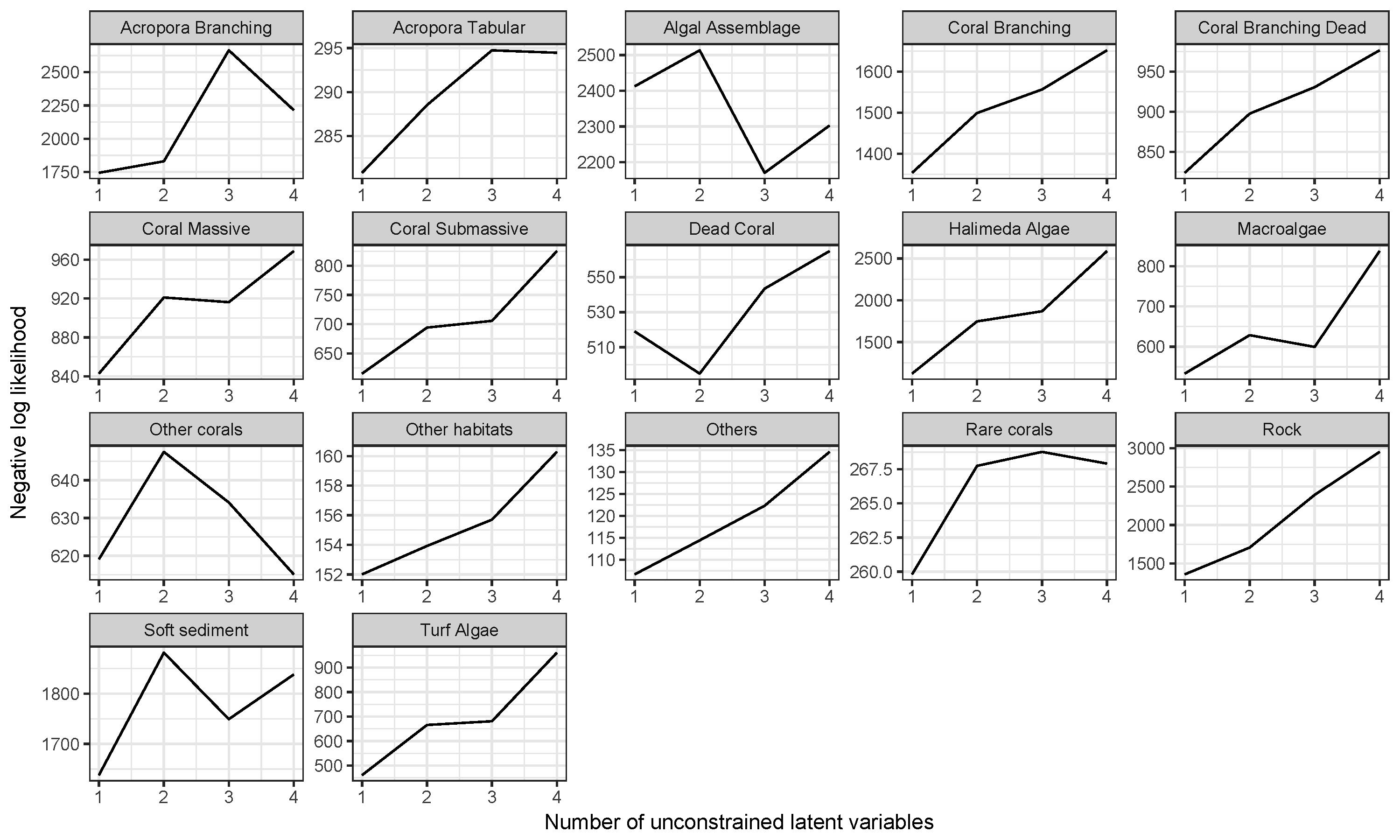


**Figure S1** Estimates for the effect of distance to log ponds (a) and high flow (b) on the constrained latent variable as a function of the number of unconstrained latent variables included in the model. Points show median estimate and bars show the 95% credibility interval.

Leave-one-out cross validation was performed on each model with 1-4 unconstrained latent variables and one constrained latent variable. To conduct the cross-validation we refit each model to 49 data-sets, each of which was missing only one data-point. We then predicted the cover of each habitat to the missing data point, using only the covariates for distance to log ponds and flow, the other unconstrained latent variables were set to zero (because they will be unknown at new sites). Thus, this was a test of the ability of the model to extrapolate to new unsampled sites where we would not have estimates for the unconstrained latent variables.

We evaluated the fit of each model by summing the negative log-likelihood of the observations given the median estimate of habitat cover (assuming a Poisson distribution). We present results per habitat, because habitats that have a stronger relationship with distance to log ponds should be better estimated than habitats that have a weak relationship with distance to log-ponds. We initialized each new MCMC chain with the median values of each parameter as estimated from the full data-set, which sped computations. We then used a burn-in of 5000 samples and took a further 5000 samples thinning every 5th sample to obtain the median estimate of habitat cover.

Results indicate that the models with 1-2 unconstrained latent variable best predicts habitat cover (Figure S2, lower values indicate a better fit) for habitats that were most strongly related to the constrained latent variable (Figure 2), which includes Branching Acropora, Tabular Acropora, Branching Coral, Halimeda Algae and Turf Algae. Habitats that were not related to the constrained latent variable, including other corals and algal assemblage had better fits when there were more unconstrained latent variables.



**Figure S2** Negative log-likelihoods for each habitat category calculated from leave-one-out cross validation. The model more closely fits the data when the negative log-likelihood is minimized.

**Table S2** Fit of the two latent variable model to each habitat type. r2 values are calculated on observed versus expected abundances. Slope values of 1 indicate no bias, <1 under-predictions and >1 over-prediction.

|  |  |  |  |
| --- | --- | --- | --- |
| Habitat | Mean abundance | r2 | Slope of expected on observed |
| Others | 0.84 | 0.19 | 0.11 |
| Other habitats | 1.86 | 0.44 | 0.32 |
| Rare corals | 2.65 | 0.20 | 0.18 |
| Acropora Tabular | 4.69 | 0.81 | 0.57 |
| Turf Algae | 4.90 | 0.37 | 0.25 |
| Macroalgae | 8.10 | 0.72 | 0.47 |
| Halimeda Algae | 11.31 | 0.96 | 0.47 |
| Algal assemblage | 12.61 | 0.96 | 0.57 |
| Submassive coral | 13.96 | 0.48 | 0.41 |
| Dead Coral | 17.39 | 0.80 | 0.74 |
| Other corals | 19.43 | 0.48 | 0.57 |
| Dead branching coral | 25.76 | 0.69 | 0.70 |
| Rock | 39.04 | 0.79 | 0.68 |
| Acropora Branching | 42.80 | 0.97 | 0.91 |
| Coral Massive | 51.12 | 0.82 | 0.85 |
| Coral Branching | 56.88 | 0.89 | 0.78 |
| Soft sediment | 61.67 | 0.94 | 0.81 |