

# Chapter 15

## Bayesian Learning of Biodiversity Models Using Repeated Observations



Ana M. M. Sequeira, M. Julian Caley, Camille Mellin,  
and Kerrie L. Mengersen

**Abstract** Predictive biodiversity distribution models (BDM) are useful for understanding the structure and functioning of ecological communities and managing them in the face of anthropogenic disturbances. In cases where their predictive performance is good, such models can help fill knowledge gaps that could only otherwise be addressed using direct observation, an often logistically and financially onerous prospect. The cornerstones of such models are environmental and spatial predictors. Typically, however, these predictors vary on different spatial and temporal scales than the biodiversity they are used to predict and are interpolated over space and time. We explore the consequences of these scale mismatches between predictors and predictions by comparing the results of BDMs built to predict fish species richness on Australia's Great Barrier Reef. Specifically, we compared a series of annual models with uninformed priors with models built using the same predictors and observations, but which accumulated information through

---

Ana M. M. Sequeira and M. Julian Caley are co-first authors.

---

**Electronic Supplementary Material** The online version of this chapter ([https://doi.org/10.1007/978-3-030-42553-1\\_15](https://doi.org/10.1007/978-3-030-42553-1_15)) contains supplementary material, which is available to authorized users.

---

A. M. Sequeira (✉)

IOMRC, UWA Oceans Institute and School of Biological Sciences, M096, The University of Western Australia, Crawley, WA, Australia

M. Caley · K. L. Mengersen

School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers, Melbourne, VIC, Australia

C. Mellin

Australian Institute of Marine Science, Townsville, QLD, Australia

The Environment Institute and School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia

time via the inclusion of informed priors calculated from previous observation years. Advantages of using informed priors in these models included (1) down-weighting the importance of a large disturbance, (2) more certain species richness predictions, (3) more consistent predictions of species richness and (4) increased certainty in parameter coefficients. Despite such advantages, further research will be required to find additional ways to improve model performance.

## 15.1 Introduction

Estimating biodiversity metrics is a central pursuit in ecological research and management. These metrics inform our understanding of the states and trends of ecosystems [13, 19], their responses to biotic and abiotic factors [11, 22, 29], and the best options for their management, conservation, and the on-going provision of ecosystem services [4, 13, 16, 18, 31].

Estimating biodiversity metrics, however, is often challenging because of the high costs of surveying and monitoring coupled with limited available resources, and because ecological systems are often highly diverse and respond in complex ways to myriad biotic and abiotic interacting factors [13]. Consequently, the data available to estimate biodiversity metrics are often insufficient to address current needs [7]. In some cases though, long-running monitoring programs provide extensive repeated measures of biological communities and can contribute to robust estimates of these metrics. Such data, however, are typically most useful for estimating the status and trends of observed biological communities, whereas estimates and their associated uncertainties are often required for entire communities across a hierarchy of spatial scales [30]. For example, the Australian Institute of Marine Science's (AIMS) Long Term Monitoring Program (LTMP) of the Great Barrier Reef (GBR) has monitored individual reefs annually for more than three decades using a spatial design that samples representative cross-shelf habitats, latitudinal sectors, and management regimes [28]. Although it is one of the most spatially extensive long-term monitoring programs on Earth, it only monitors a small fraction of all the reefs present (<2% of all GBR reefs). Consequently, biodiversity estimation beyond this relatively small set of reefs must rely on predictions for unmonitored reefs (e.g. [15]) and requires good predictability into unsampled space (a component of spatial statistical modeling: [25]).

Where it is desirable to predict biodiversity into a larger domain than a series of observed communities, biodiversity distribution models (BDM), a general case of species distribution models (SDM) where the response variable may be a composite metric such as species richness or total abundance across species, can be constructed using combinations of environmental and spatial variables. These models can then be useful to predict biodiversity metrics across domains where values are not observed [15, 24, 26, 32, 33]. While such models have proven effective to varying degrees, there is commonly a mismatch between the states and dynamics of the ecological communities being predicted and the environmental and spatial observations

used to predict them. For example, ecological communities are typically sampled at regular intervals (e.g., yearly). Spatial predictors, such as a community's location relative to geological features, vary over geological time and can be assumed to be invariant with respect to ecological prediction, whereas environmental predictors vary on a variety of temporal and spatial scales. For the purposes of predictive modelling, these environmental predictors are often available only as long-term annual averages and spatially interpolated to common scales (e.g. [marinehub.org](http://marinehub.org)). Consequently, repeated measures of ecological metrics often rely on a set of predictors that are inherently less variable than the metric they are being used to predict, either because of the characteristics of the predictor (e.g. spatial predictors) or the way it was collected and processed (e.g. environmental predictors). These characteristics of predictors may in turn compromise the performance of predictive models given that a diversity metric might vary through time and space at rates unrelated to the variables used to predict it.

The consequences of these scale mismatches between predictors and predictions on the performance of such models are likely to vary from year to year, as the response variable changes but the predictor variables do not. These changes have the potential to affect the predictive performance of a model in a number of ways including the ability to predict true values and their associated uncertainties, the coefficients of the predictors estimated for the model, and the structure of the best performing model. Despite the potential importance of such mismatches, understanding of their influence on the construction and application of BDMs is poor. To begin addressing this knowledge gap, we explore ways in which analytical approaches to building and applying BDMs affect their predictive performance, the estimation of the coefficients of model parameters, and the selection of the best model structure in cases where recorded values of the response variables vary in space or time but observations of predictor variables vary less over time.

When repeated observations from a monitoring program are available, common approaches to their analysis include considering each repeated set of observations separately and then making post-hoc comparisons between them to infer community states through time, or averaging all data across replicates and then estimating the best model. A disadvantage of such approaches is that they fail to use all the information available from repeated observations to help understand how a predictive model might improve as monitoring continues through time. Comparisons between models can also be difficult as the importance of predictors change between years. Moreover, understanding how such information accumulates through time can facilitate more efficient and effective allocation of limited and valuable monitoring resources through the implementation of adaptive sampling designs [12]. By adopting a Bayesian learning approach to this problem, it should be possible to better understand how the performance of such models changes with the addition of information through time. In such an approach, the results obtained from a previous survey or surveys can be used as prior information in the analysis of the data for the latest survey. Adopting this approach results in an iterative updating of information as it becomes available which has theoretical and computational advantages. Theoretically, compared with the independent analysis

of surveys described above, the obtained estimates should move closer to true values more quickly and smoothly, any trend in the replicate estimates should be smoother, the estimates obtained in each replicate should be more precise (i.e., have narrower credible intervals), and post-hoc analysis across the time series should no longer be needed. Similarly, the estimates of the coefficients of the model and the model structure should converge to the true values and be associated with progressively decreasing uncertainties. Computationally, compared with analysing all available data each time new observations become available, the Bayesian learning approach does not require reanalysis of the entire dataset during updating but instead requires a simpler and less computationally costly analysis of the current data and the prior which encapsulates information from past surveys.

Bayesian modelling of data with repeated measures is now commonplace and offers advantages over other approaches in terms of estimation, model flexibility, and inference [3]. Bayesian learning, also known as recursive Bayesian estimation or Bayesian filtering, is commonly employed for a wide variety of problems that require iterative updating of information from quality monitoring and control [1] to analyses of streaming data [23]. To explore the comparative benefits of using a Bayesian learning approach for estimating biodiversity using typical monitoring data, we analysed species richness patterns of fishes on Australia's Great Barrier Reef (GBR). We analyse the annual LTMP data using uninformed priors for each year analogous to the frequentist analyses of individual repeated observations, and compare these results to those obtained using informed priors derived from previous observations. Based on the results of these analyses we make recommendations for improved learning where a set of temporally less variable predictors are used to make predictions from observations that vary to a greater extent through time.

## 15.2 Methods

### 15.2.1 *Fish Species Counts and Environmental and Spatial Predictors*

We used counts of fish species on the GBR collected by the Australian Institute of Marine Science's (AIMS) Long-Term Monitoring Program (LTMP) [28] for the years 2003–2013. For this period, annual survey data were available for each year from 2003 to 2005 and every second year after 2005 due to a change of sampling design. A total of 46 reefs were monitored across six latitudinal sectors (Cooktown-Lizard Island, Cairns, Townsville, Whitsunday, Swain and Capricorn-Bunker) spanning 150,000 km<sup>2</sup> of the GBR. In each sector, with the exception of the Swain and Capricorn-Bunker sectors, at least two reefs were sampled in each of three shelf positions (i.e., inner, mid- and outer). At each reef, 5 transects in each of 3 sites were sampled and we analysed observations from the same 133 locations in each of these 7 years. Observations were made using transect-based underwater visual survey. Transects were randomly selected, permanently marked, and ran roughly parallel to

the reef crest, each separated by at least 10 m along the 6–9 m depth contour. Counts of 251 fish species from across 10 taxonomic families were recorded. This set of species excluded cryptic and nocturnal species. Larger mobile species were counted first along a 5 m wide transect, and smaller, less mobile species (e.g. damselfishes: Pomacentridae) were counted in a 1-m wide strip along the same transect during the return swim (for detailed methods and species counted, see [10]). To prevent potential systematic bias in the fish counts associated with different observers, calibration of all divers occurred annually [10]. To predict these fish species counts (i.e. species richness), we used both environmental and spatial predictors. We used a set of environmental predictors available for Australia at a national scale and at a  $0.01^\circ$  resolution ([marinehub.org](http://marinehub.org)) including sea surface temperature (SST), chlorophyll-*a* (Chl *a*), salinity, nutrients (NO<sub>3</sub>, PO<sub>4</sub>, and SI), light (K490av), depth (as a proxy for habitat), oxygen, and sediment characteristics including percentages of carbonates, gravel, sand, and mud [14]. To account for geographical effects on the distributional patterns of reef fishes, we also included two spatial predictors: the shortest distances to coast (*coast*) and to the outer limit of the reefs (*barrier*), which have been used to successfully predict fish species richness and abundances on the GBR [15, 24]. We calculated these distances for each sampled site and node on the  $0.01^\circ$  national grid using the Near tool in ArcGIS10.1 (ESRI, Redlands, CA, USA) and an equidistant cylindrical coordinate system. We then assigned each sampling site to the closest node on the  $0.01^\circ$  national grid and used the environmental and spatial predictors corresponding to these locations.

### 15.2.2 Bayesian Models

Using *reef* as a random effect to account for the hierarchical nature of the dataset with sites nested within reefs, we developed Bayesian generalized linear mixed-effects models (GLMM) of fish species richness assuming a Poisson distributed response  $S_{ij}$  for the  $i$ th location in the  $j$ th reef, with a log-link and linear and quadratic regression terms for the covariates  $X_{ij}$ . Seven separate models were developed using each yearly dataset of fish species richness observations from the GBR as a response variable (Table 15.1). Allowing for extra-Poisson variation through a residual  $\varepsilon_i \sim N(0, \sigma^2)$ , the likelihood is thus given by

$$S_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\log(\mu_{ij}) = \alpha_j + X\beta + \varepsilon_i.$$

An uninformative Gaussian prior (i.e., zero mean and relatively large variance) was specified for the random effect for reef,  $\alpha_j$ . Twelve combinations of covariates were considered for each set of yearly models (Table 15.1). Univariate priors for each of the regression coefficients in the vector  $\beta$  were specified. Two sets of such priors were considered. First, we used independent uninformative Gaussian priors

**Table 15.1** Description of fitted models

No.	Model description	Covariates included in model
1	Comprehensive model	Reef + Coast + Coast <sup>2</sup> + Barrier + Barrier <sup>2</sup> + Depth + Depth <sup>2</sup> + Slope + O <sub>2</sub> + SST + SST <sup>2</sup> + Light
2	Distance to domain boundaries	Reef + Coast + Coast <sup>2</sup> + Barrier + Barrier <sup>2</sup>
3	Physical predictors	Reef + Depth + Depth <sup>2</sup> + Slope + Aspect
4	Particular sediment type	Reef + Gravel
5	Particular sediment type	Reef + Sand
6	Particular sediment type	Reef + Mud
7	Nutrients	Reef + NO <sub>3</sub> + PO <sub>4</sub> + Silica
8	Oxygen and salinity	Reef + O <sub>2</sub> + Salinity
9	Productivity	Reef + Chl <i>a</i>
10	Temperature	Reef + SST + SST <sup>2</sup>
11	Light availability	Reef + Light
12	Intercept only	Reef

All models included a random effect for reef. Coast: distance to coast; Barrier: distance to the reef's outer limit; Gravel, Sand and Mud represent percentage of gravel, sand, and mud, respectively; average concentrations of NO<sub>3</sub>: nitrate, PO<sub>4</sub>: phosphate, SI: silicate, O<sub>2</sub>: dissolved oxygen, Chl *a*: chlorophyll *a*; Sal: salinity; SST: average annual sea surface temperature; Light: coefficient of light attenuation at 490 nm. All predictors were mean centred, and superscript 2 indicates predictors included as quadratic terms. Sediment variables (gravel, mud and sand) were included in separate models (4–6) due to collinearity

as above, assuming no prior knowledge of the relationships between the response variable and the set of predictors being included in each model. This approach provided a baseline against which we compared a second set of models using informed priors and which therefore could exploit potential benefits of Bayesian methods. In this second approach, we modelled the first year's observations using uninformative Gaussian priors as described above. Consequently, the results of both approaches will be the same for the first year. In each subsequent year, we used the posterior mean and variance from the previous year(s) to construct an informed Gaussian prior, which was then used to model the responses of the current year of observation (Table S2).

We used a Markov Chain Monte Carlo (MCMC) algorithm with 100,000 iterations, a burn in of 10,000, a thinning rate of 2 (i.e., discarding every second simulated value to reduce autocorrelation and Monte Carlo error), and ran three chains to check convergence. To ensure the behaviour of the chains would not differ for larger MCMC runs, we also compared results from 600,000 iterations after burn in. Due to limits to computational power, we ran this larger MCMC in steps of 20,000 iterations by updating the chains with the last value obtained in each of the previous iterations. The modelling results shown here are derived from the last 10,000 iterations in each MCMC, having ensured convergence had been reached based on the Gelman-Brooks-Rubin diagnostic (i.e., *rhat* < 1.1). The retained MCMC samples were used to obtain posterior means and 95% credible intervals (CrI). CrIs were estimated for parameters of interest and a posterior

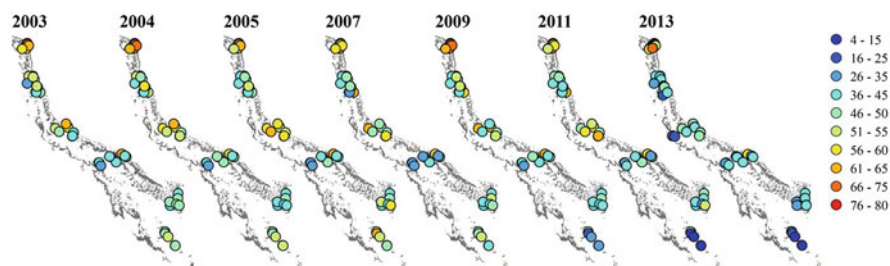
predictive check of their individual contributions made using the sum of squared Pearson residuals, the raw residual divided by the square root of the variance.

To understand the effects of the different modelling methods on our results, we used wBIC and wAICc for comparison, as the use of DIC and wDIC can be inconsistent for GLMM [17]. Moreover, the wAICc diagnostic provided a more straightforward comparison with previous published results obtained using a frequentistic approach (e.g. [15]). We also included a posterior predictive check and report Bayesian  $p$ -values to assess the resulting predictions from our Bayesian models. To predict species richness across the entire GBR, we used a model-averaging procedure using wAICc to average the set of model formulations included in each model run.

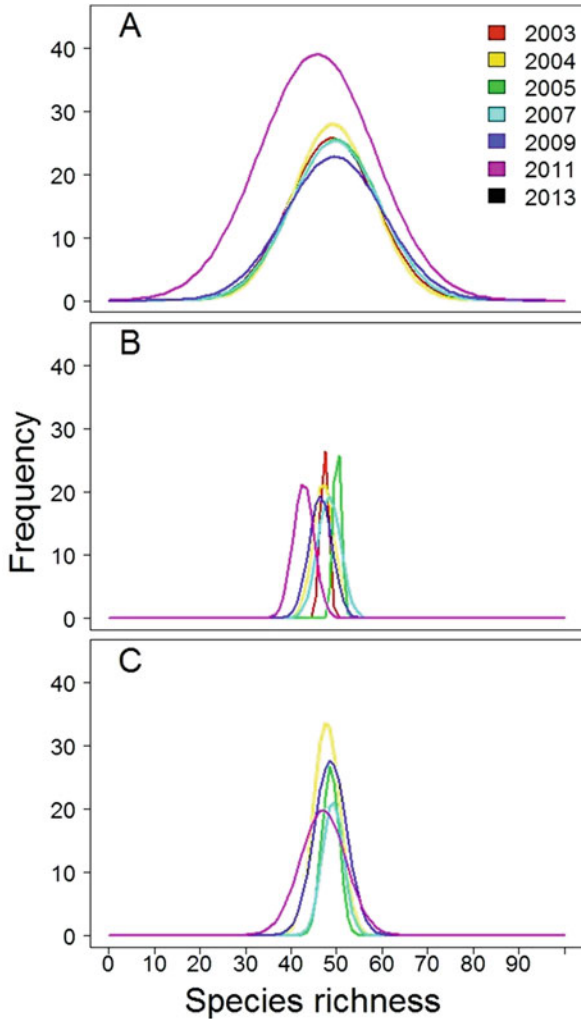
### 15.3 Results

Observed species richness of fishes varied among years with the greatest species richness densities and variation among reefs recorded in 2011 (range: 10–80 species). The peak density of this year was also shifted left compared to the other years of observations (Figs. 15.1 and 15.2a), which displayed less variability (range: 25–70 species) and lower peak densities (Figs. 15.1 and 15.2a). With independent priors, the Bayesian models identified Chl  $a$ , SST, or light as influential in models 9, 10 and 11, respectively, for all yearly datasets, with emphasis differing between linear and quadratic terms for SST in different years. PO<sub>4</sub> (model 7) in years 2005, 2007 and 2013, and salinity (model 8) in all years also showed substantively non-zero effects in that the 95% CrIs excluded zero. The analogous CrIs for the coefficients of all other predictors overlapped zero (Table S1). Results for models and datasets using informative priors were similar but with some additional effects observed for predictors included in models 4–7 in early years only (Table S2).

Chl  $a$  and Light in models 9 and 11, respectively, were the only two predictors for which the 95% CrIs for the coefficient estimates excluded zero for all



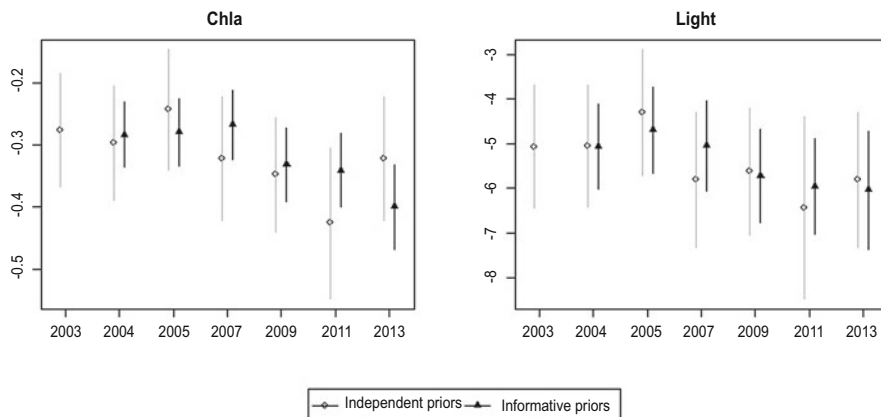
**Fig. 15.1** Map of sampled species richness in the GBR across six latitudinal sections of the Great Barrier Reef (GBR) and three shelf positions (outer, mid and inner). Legend indicates number of species recorded per site after pooling counts made using  $5 \times 50$ -m long transects per site



**Fig. 15.2** Density plot of observed (a) and predicted species richness across datasets when using independent (b) or informative priors (c)

datasets irrespective of the application of informative or non-informative priors. The coefficient estimates for these two predictors averaged across models within years tended to be more negative toward the end of this time series (Fig. 15.3). Models with uninformative priors varied in goodness of fit according to wBIC, but generally models 2 (reef and reef position relative to spatial domains), 10 (reef and temperature), and 11 (reef and light) were among the best-fitting models (Table 15.2). As expected, when using informative priors, wBIC values were generally





**Fig. 15.3** Estimates of coefficient for Chl *a* and Light. Results are shown for each model run across datasets and when using independent (open circle with grey standard deviation lines) and informative priors (filled triangle with black standard deviation line)

lower. Goodness of fit according to wAICc showed similar patterns to those obtained with wBIC (Table 15.2).

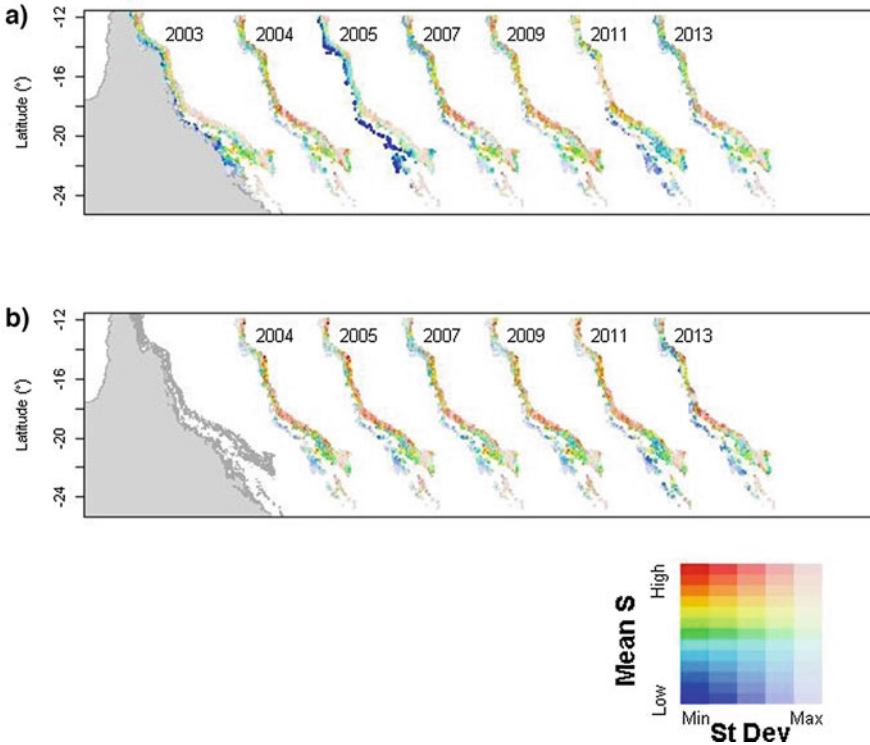
### 15.3.1 Model Predictions

Density plots of species richness predictions demonstrate that each model set differed across datasets, but were more consistent when using informative priors (cf. Fig. 15.2b, c). Posterior predictive checks resulted in Bayesian  $p$ -values close to 0.5, indicating good predictive performance only for years 2011 and 2009, ranging respectively from 0.349–0.511 and 0.605–0.682 for uninformed priors and from 0.354–0.503 and 0.625–0.689 for informed priors. For all other datasets across all model runs, the Bayesian  $p$ -value was always close to one (0.916–0.989 for uninformed, and 0.916–0.989 for informative priors), indicating poor fits between observed and predicted species richness distribution, and hence, poor predictive performance. In no case, however, was the Bayes  $p$ -value  $>0.99$ , which would indicate major failure of model fit [8]. Irrespective of the use of independent or informative priors, higher fish species richness was predicted mostly in the northern-central offshore reefs, with the difference between inner and outer reefs being more marked in some years (e.g., 2003, 2005 and 2011 with independent priors) (Fig. 15.4).

**Table 15.2** wBIC and wAICc calculated for each model and for all datasets considered when using independent or informative priors

Model	wBIC												wAICc											
	2003	2004	2005	2007	2009	2011	2013	2003	2004	2005	2007	2009	2011	2013	2003	2004	2005	2007	2009	2011	2013			
Independent priors	<b>1</b>	0.006	0.001	0.003	-	0.018	0.073	-	0.011	0.003	0.005	0.001	0.024	<b>0.104</b>	0.017	<b>0.146</b>	0.017	-	-	-	-	-		
	<b>2</b>	<b>0.289</b>	0.037	<b>0.676</b>	<b>0.132</b>	<b>0.150</b>	0.012	<b>0.132</b>	<b>0.263</b>	0.052	<b>0.579</b>	<b>0.146</b>	<b>0.168</b>	0.017	<b>0.146</b>	0.017	<b>0.146</b>	0.017	-	-	-	-		
	<b>3</b>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	<b>4</b>	0.002	0.004	0.004	0.009	0.006	0.025	0.009	0.006	0.009	0.009	0.009	0.015	0.033	0.015	0.033	0.015	0.033	0.015	0.033	0.015	0.033		
	<b>5</b>	0.002	0.008	0.007	0.010	0.006	0.014	0.010	0.005	0.017	0.013	0.016	0.012	0.020	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016		
	<b>6</b>	0.003	0.012	0.009	0.018	0.007	0.022	0.018	0.008	0.025	0.016	0.027	0.014	0.029	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027		
	<b>7</b>	0.001	-	0.003	0.002	-	-	0.002	0.002	0.001	0.008	0.003	0.001	-	0.003	0.001	0.003	0.001	-	-	-	0.003		
	<b>8</b>	0.009	0.023	0.020	0.059	0.016	0.015	0.059	0.016	0.030	0.033	0.084	0.023	0.016	0.084	0.023	0.084	0.023	0.016	0.084	0.023	0.084		
	<b>9</b>	0.022	<b>0.105</b>	0.013	0.058	<b>0.176</b>	<b>0.101</b>	0.058	0.030	<b>0.108</b>	0.018	0.055	<b>0.159</b>	<b>0.108</b>	0.055	<b>0.159</b>	<b>0.108</b>	0.055	0.018	<b>0.108</b>	0.055	0.055		
	<b>10</b>	<b>0.466</b>	<b>0.279</b>	<b>0.196</b>	<b>0.193</b>	<b>0.147</b>	<b>0.543</b>	<b>0.193</b>	<b>0.446</b>	<b>0.300</b>	<b>0.234</b>	<b>0.217</b>	<b>0.171</b>	<b>0.457</b>	<b>0.217</b>	<b>0.171</b>	<b>0.457</b>	<b>0.217</b>	<b>0.171</b>	<b>0.457</b>	<b>0.217</b>	<b>0.171</b>		
	<b>11</b>	<b>0.196</b>	<b>0.523</b>	0.060	<b>0.502</b>	<b>0.464</b>	<b>0.148</b>	<b>0.502</b>	<b>0.205</b>	<b>0.440</b>	0.069	<b>0.408</b>	<b>0.397</b>	<b>0.160</b>	<b>0.408</b>	<b>0.397</b>	<b>0.160</b>	<b>0.408</b>	<b>0.397</b>	<b>0.160</b>	<b>0.408</b>	<b>0.397</b>		
	<b>12</b>	0.004	0.008	0.009	0.018	0.010	0.047	0.018	0.010	0.016	0.016	0.027	0.019	0.058	0.027	0.019	0.058	0.027	0.019	0.058	0.027	0.019		
Informative priors	<b>1</b>	-	<b>0.393</b>	<b>0.347</b>	<b>0.274</b>	<b>0.529</b>	<b>0.975</b>	0.038	-	<b>0.328</b>	<b>0.264</b>	<b>0.228</b>	<b>0.450</b>	<b>0.975</b>	<b>0.450</b>	<b>0.228</b>	<b>0.264</b>	<b>0.228</b>	<b>0.450</b>	<b>0.975</b>	0.038			
	<b>2</b>	-	0.073	<b>0.307</b>	<b>0.225</b>	<b>0.110</b>	0.001	0.080	-	0.084	<b>0.320</b>	<b>0.203</b>	0.000	0.089	0.000	0.089	0.000	0.089	0.000	0.089	0.000			
	<b>3</b>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	<b>4</b>	-	0.001	0.002	0.004	0.002	-	0.021	-	0.004	0.004	0.008	0.006	0.001	0.008	0.006	0.001	0.008	0.006	0.001	0.008	0.006		
	<b>5</b>	-	0.003	0.003	0.009	0.003	-	0.029	-	0.008	0.007	0.014	0.008	0.001	0.008	0.001	0.008	0.007	0.014	0.008	0.001	0.008		
	<b>6</b>	-	0.003	0.004	0.009	0.005	-	0.021	-	0.008	0.008	0.016	0.011	0.001	0.008	0.011	0.001	0.008	0.016	0.011	0.001	0.008		
	<b>7</b>	-	-	0.006	0.004	-	-	0.003	-	0.001	0.015	0.008	0.001	-	0.006	0.001	-	0.008	0.001	-	0.006	0.001		
	<b>8</b>	-	0.009	0.027	0.076	0.010	0.001	<b>0.224</b>	-	0.014	0.042	<b>0.106</b>	0.016	0.002	<b>0.238</b>	0.016	0.002	<b>0.106</b>	0.016	0.002	<b>0.238</b>	0.016		
	<b>9</b>	-	0.029	0.015	0.027	0.063	0.009	0.048	-	0.040	0.018	0.032	0.071	0.014	0.042	0.071	0.014	0.032	0.071	0.014	0.042	0.071		
	<b>10</b>	-	<b>0.313</b>	<b>0.225</b>	<b>0.128</b>	<b>0.128</b>	0.062	<b>0.193</b>	-	<b>0.323</b>	<b>0.250</b>	<b>0.147</b>	0.087	0.011	<b>0.188</b>	0.087	0.011	<b>0.147</b>	0.087	0.011	<b>0.188</b>	0.087		
	<b>11</b>	-	<b>0.172</b>	0.061	<b>0.240</b>	<b>0.212</b>	0.005	<b>0.318</b>	-	<b>0.183</b>	0.064	<b>0.228</b>	<b>0.207</b>	0.009	<b>0.276</b>	<b>0.207</b>	0.009	<b>0.228</b>	<b>0.207</b>	0.009	<b>0.276</b>	0.009		
	<b>12</b>	-	0.002	0.003	0.005	0.003	0.001	0.026	-	0.006	0.007	0.009	0.008	0.001	0.006	0.008	0.001	0.009	0.008	0.001	0.006	0.008		

Shown only for wBIC and wAICc > 0.001, with bold indicating values > 0.01. Model descriptions are given in Table 15.1



**Fig. 15.4** Model predictions of species richness across the entire GBR for each model set run from yearly datasets when using independent (a) and informed (b) priors. Figure shows model-averaged results when using wAICc to average the contribution of each model in the model set

## 15.4 Discussion

Despite the difficulty and expense of observing complex natural ecosystems [13], the need to estimate ecosystem states and trajectories through time as they are influenced by increasingly frequent and severe disturbances is becoming more urgent. It is important, therefore, to understand how best to use information currently available to estimate these states and trajectories, understand their causes, and how to optimize the design of survey and monitoring programs to improve our understanding of these dynamics. In light of these information needs, we have compared here a series of independent annual analyses using uninformative priors with a recursive approach using informative priors based on previous data. The models were evaluated in the context of constructing and testing BDMs with specific reference to the prediction of fish species richness on Australia's Great Barrier Reef. The performances of the models, constructed using either of these approaches, indicate further room for improvement in how such models are constructed and

some advantages to the use of informed priors. We detail below knowledge gained that would not have been possible using uninformed priors alone.

It is widely appreciated that natural ecosystems are affected by a variety of disturbances that can have large effects on the states and subsequent trajectories of the biological communities they host. Many direct observations of such impacts and recovery are now recorded in the literature (e.g., [5, 6, 9, 16, 20, 21, 27]). Such disturbances, however, can also affect model selection of BDMs by affecting the inclusions of particular predictors that are upweighted in specific situations where extreme values are reached because of either the immediate or longer-term effects of disturbances. Consequently, over the longer term, it may be desirable to down-weight or average these effects to achieve a more general view of the role of these predictors. For example, in early 2011 cyclone Yasi, one of the largest Australian cyclones over the past 20 years, caused extensive damage to the coral communities of the GBR [2]. Yasi also seems to have affected the species richness of the fish communities both in terms of the densities of species observed and their variability (Fig. 15.2a). Our study suggests that the impacts from such rare events on predicted species richness was much less for the models with informed priors that showed much greater consistency in both density and variation across all modeled years.

Bayesian approaches also provide opportunities to assess estimates and uncertainties in parameter values and model structure. In our study, Chl *a* and Light, were the only two predictors for which coefficient estimates were substantively different from zero for all datasets, irrespective of the application of informative or non-informative priors. These two parameters, therefore, provide an opportunity to examine the effect of these two modelling approaches on estimating their values and uncertainties across years as the Bayesian priors contained progressively more information. In all cases, estimates based on informed priors were more certain, however, modal values were not consistently greater or smaller, nor were the credible intervals progressively smaller as information accumulated in the informed priors and these intervals overlapped extensively. The two modelling frameworks also nominated different model structures as best. The recursive analysis, using informed priors based on previous survey data, indicated statistical contributions from more predictors than did the independent analysis of each time period. Accordingly, more information was harnessed by using informed priors indicating that greater investment in observations of these predictors may have additional utility.

While much was learned here by comparing Bayesian models with informed and uninformed priors, neither model performed very well with respect to prediction, indicating much is still to be learned regarding how best to increase their performance. The options for improvement here are many and will depend on the interests and opportunities of individual researchers and their groups. In contrast to this study, previous studies of predictive models based on these and similar data were better able to predict species richness and abundance by averaging these responses across the time-series of observations [15, 24]. Therefore, the challenge of better predictive performance identified here appears to be in generating finer-scale temporal predictions. Where predictions at this scale are desirable, our results

suggest Bayesian models with informed priors can be useful for better selection of model structures, estimation of their parameters, and the down weighting of rare but significant events. Nonetheless, even though the time series used here to build these predictive models was spatially extensive and long compared to many ecological data series, the complexity of the processes, and potentially their non-stationarity, that can configure a metric such as species richness are likely to remain challenging. This challenge is likely to be exacerbated when predicting other biodiversity metrics such as abundances of individual species or abundances summed across species.

It is also clear that much longer time series may be required before prior probabilities can become sufficiently informed to facilitate more substantial reductions in parameter uncertainty. In the meantime, however, even a modest number of repeated observations appears to inform priors sufficiently to obtain quite consistent 95% CrIs and modal predicted values.

## References

1. H. Assareh, I. Smith, K.L. Mengersen, Bayesian change point detection in monitoring clinical outcomes, in *Case Studies in Bayesian Statistical Modelling and Analysis*, ed. by C. L. Alston, K. L. Mengersen, A. N. Pettitt, (Wiley, Chichester, 2012)
2. R. Beeden, J. Maynard, M. Puotinen, P. Marshall, J. Dryden, J. Goldberg, G. Williams, Impacts and recovery from severe tropical cyclone Yasi on the Great Barrier Reef. *PLoS One* **10**, e0121272 (2015)
3. L. Broemeling, *Bayesian Methods for Repeated Measures* (Chapman and Hall/CRC, New York, 2016)
4. B.J. Cardinale, J.E. Duffy, A. Gonzalez, D.U. Hooper, C. Perrings, P. Venail, A. Narwani, G.M. Mace, D. Tilman, D.A. Wardle, A.P. Kinzig, G.C. Daily, M. Loreau, J.B. Grace, A. Larigauderie, D.S. Srivastava, S. Naeem, Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012)
5. J.H. Connell, Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302–1310 (1978)
6. P.K. Dayton, Competition, disturbance, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. *Ecol. Monogr.* **41**, 351–389 (1971)
7. R. Fisher, B.T. Radford, N. Knowlton, R.E. Brainard, F.B. Michaelis, M.J. Caley, Global mismatch between research effort and conservation needs of tropical coral reefs. *Conserv. Lett.* **4**, 64–72 (2011)
8. A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin, *Bayesian Data Analysis* (Chapman and Hall, New York, 2015)
9. A.R. Halford, M.J. Caley, Towards an understanding of resilience in isolated coral reefs. *Glob. Chang. Biol.* **15**, 3031–3045 (2009)
10. A.R. Halford, A.A. Thompson, *Visual Census Surveys of Reef Fish. Long Term Monitoring of the Great Barrier Reef Standard Operational Procedure Number 3* (Australian Institute of Marine Science, Townsville, 1996)
11. A.R. Ives, S.R. Carpenter, Stability and diversity of ecosystems. *Science* **317**, 58–62 (2007)
12. S.Y. Kang, J.M. McGree, C.C. Drovandi, M.J. Caley, K.L. Mengersen, Bayesian adaptive design: improving the effectiveness of monitoring of the Great Barrier Reef. *Ecol. Appl.* **26**, 2635–2646 (2016)
13. G.M. Lovett, D.A. Burns, C.T. Driscoll, J.C. Jenkins, M.J. Mitchell, L. Rustad, J.B. Shanley, G.E. Likens, R. Haeuber, Who needs environmental monitoring? *Front. Ecol. Environ.* **5**, 253–260 (2007)

14. S.A. Matthews, C. Mellin, A. MacNeil, S.F. Heron, W. Skirving, M. Puotinen, M.J. Devlin, M. Pratchett, High-resolution characterization of the abiotic environment and disturbance regimes on the Great Barrier Reef, 1985–2017. *Ecology* **100**, e02574 (2019)
15. C. Mellin, C.J.A. Bradshaw, M.G. Meekan, M.J. Caley, Environmental and spatial predictors of species richness and abundance in coral reef fishes. *Glob. Ecol. Biogeogr.* **19**, 212–222 (2010)
16. C. Mellin, M. Aaron MacNeil, A.J. Cheal, M.J. Emslie, M. Julian Caley, Marine protected areas increase resilience among coral reef communities. *Ecol. Lett.* **19**, 629–637 (2016)
17. R.B. Millar, Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics* **65**, 962–969 (2009)
18. N. Myers, R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000)
19. J.D. Nichols, B.K. Williams, Monitoring for conservation. *Trends Ecol. Evol.* **21**, 668–673 (2006)
20. R.T. Paine, S.A. Levin, Intertidal landscapes: disturbance and the dynamics of pattern. *Ecol. Monogr.* **51**, 145–178 (1981)
21. S.T.A. Pickett, P.S.E. White, *The Ecology of Natural Disturbance and Patch Dynamics* (Academic, London, 1985)
22. S.L. Pimm, The complexity and stability of ecosystems. *Nature* **307**, 321–326 (1984)
23. S. Sarkka, *Bayesian Filtering and Smoothing* (Cambridge University Press, Cambridge, 2013)
24. A.M.M. Sequeira, C. Mellin, H.M. Lozano-Montes, M.A. Vanderklift, R.C. Babcock, M.D.E. Haywood, J.J. Meeuwig, M.J. Caley, Transferability of predictive models of coral reef fish species richness. *J. Appl. Ecol.* **53**, 64–72 (2016)
25. A.M.M. Sequeira, P.J. Bouchet, K.L. Yates, K. Mengersen, M.J. Caley, Transferring biodiversity models for conservation: opportunities and challenges. *Methods Ecol. Evol.* **9**, 1250–1264 (2018)
26. A.M.M. Sequeira, C. Mellin, H.M. Lozano-Montes, J.J. Meeuwig, M.A. Vanderklift, M.D.E. Haywood, R.C. Babcock, M.J. Caley, Challenges of transferring models of fish abundance between coral reefs. *PeerJ* **6**, e4566 (2018)
27. W.P. Sousa, The role of disturbance in natural communities. *Annu. Rev. Ecol. Syst.* **15**, 353–391 (1984)
28. H. Sweatman, A. Cheal, G. Coleman, M. Emslie, K. Johns, M. Jonker, I. Miller, K. Osborne, *Long-Term Monitoring of the Great Barrier Reef. Status Report No 8* (Australian Institute of Marine Science, Townsville, 2008)
29. D. Tilman, J.A. Downing, Biodiversity and stability in grasslands. *Nature* **367**, 363 (1994)
30. J. Vercelloni, K. Mengersen, F. Ruggeri, M.J. Caley, Improved coral population estimation reveals trends at multiple scales on Australia's Great Barrier Reef. *Ecosystems* **20**, 1337–1350 (2017)
31. B. Worm, E.B. Barbier, N. Beaumont, J.E. Duffy, C. Folke, B.S. Halpern, J.B.C. Jackson, H.K. Lotze, F. Micheli, S.R. Palumbi, E. Sala, K.A. Selkoe, J.J. Stachowicz, R. Watson, Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**, 787–790 (2006)
32. K.L. Yates, C. Mellin, M.J. Caley, B.T. Radford, J.J. Meeuwig, Models of marine fish biodiversity: assessing predictors from three habitat classification schemes. *PLoS One* **11**, e0155634 (2016)
33. K.L. Yates, P.J. Bouchet, M.J. Caley, K. Mengersen, C.F. Randin, S. Parnell, A.H. Fielding, A.J. Bamford, S. Ban, A.M. Barbosa, C.F. Dormann, J. Elith, C.B. Embling, G.N. Ervin, R. Fisher, S. Gould, R.F. Graf, E.J. Gregor, P.N. Halpin, R.K. Heikkinen, S. Heinänen, A.R. Jones, P.K. Krishnakumar, V. Lauria, H. Lozano-Montes, L. Mannocci, C. Mellin, M.B. Mesgaran, E. Moreno-Amat, S. Mormede, E. Novaczek, S. Opiel, G. Ortuño Crespo, A.T. Peterson, G. Rapacciolo, J.J. Roberts, R.E. Ross, K.L. Scales, D. Schoeman, P. Snelgrove, G. Sundblad, W. Thuiller, L.G. Torres, H. Verbruggen, L. Wang, S. Wenger, M.J. Whittingham, Y. Zharikov, D. Zurell, A.M.M. Sequeira, Outstanding challenges in the transferability of ecological models. *Trends Ecol. Evol.* **33**, 790–802 (2018)