CONTRIBUTED PAPER

Conservation Science and Practice
A journal of the Society for Conservation Biology

WILEY

Transferable, predictive models of benthic communities informs marine spatial planning in a remote and data-poor region

Tom C. L. Bridge[1,2] | Zhi Huang[3] | Rachel Przeslawski[3] | Maggie Tran[3] | Justy Siwabessy[3] | Kim Picard[3] | April E. Reside[4] | Murray Logan[5] | Scott L. Nichol[3] | M. Julian Caley[6]

[1]Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia

[2]Biodiversity and Geosciences Program, Museum of Tropical Queensland, Queensland Museum Network, Townsville, Queensland, Australia

[3]Geoscience Australia, National Earth and Marine Observations Branch, Canberra, Australian Capital Territory, Australia

[4]Centre for Biodiversity and Conservation Science, School of Biological Sciences, The University of Queensland, St Lucia, Queensland, Australia

[5]Australian Institute of Marine Science, Townsville, Queensland, Australia

[6]ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Brisbane, Queensland, Australia

Correspondence
Tom C. L. Bridge, Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, 101 Angus Smith Drive, Townsville, QLD 4811, Australia.
Email: thomas.bridge@jcu.edu.au

Abstract

Systematic conservation planning requires spatial information on biodiversity. Such information is often unavailable, forcing spatial planning to rely on assumed relationships between species and environmental features. This problem is particularly acute in large, remote marine protected areas that are proliferating rapidly. Here, we use models to predict whether (a) macrobenthic biodiversity across four taxa (gorgonians, soft corals, hard corals, and sponges) with different life histories are congruent within seascape features through regional space; and (b) models generated in an intensively-sampled area in one region can predict the occurrence of habitat-forming macrobenthos in neighboring ones. All four taxa studied showed similar habitat preferences, but high variability in distributions among and within features suggesting factors other than simple geomorphology influence these regional biodiversity patterns. Nonetheless, models derived from one region accurately predicted the presence and absence of the same taxa hundreds of kilometers away. This transferability of models of species occurrences has the potential to deliver improved reserve design in data-deficient regions.

KEYWORDS

boosted regression trees, conservation planning, marine protected areas, predictive models, spatial planning, transferability

1 | INTRODUCTION

Systematic conservation planning requires data on the spatial distributions of species among regions (Margules & Pressey, 2000). However, such spatial data are often incomplete, particularly in remote areas, where data collection can be difficult because of high costs of data acquisition and other logistical challenges. Consequently, broad-scale habitat types, or taxonomic groups are often used as surrogates of biodiversity (Lombard, Cowling, Pressey, & Rebelo, 2003; Mellin et al., 2011; Sutcliffe, Mellin, Pitcher, Possingham, & Caley, 2014). In such cases, the effectiveness of conservation actions depends on the ability of such surrogates to adequately represent biodiversity patterns. However, the effectiveness of these surrogates can vary considerably depending on a range of factors including the taxon or spatial scale examined (Grantham, Pressey, Wells, & Beattie, 2010; Lombard et al., 2003; Mellin et al., 2011; Sutcliffe, Pitcher, Caley, & Possinghan, 2012), the nature of the surrogate (e.g., richness, vs. abundance vs. biomass, Yates, Mellin, Caley, Radford, & Meeuwig, 2016), the predictors used and how they are weighted (Mellin, Mengersen, Bradshaw, & Caley, 2014), the method of data collection (e.g., remotely sensed vs. observer classified, Yates et al., 2016), and the nature and extent of the extrapolation required to predict biodiversity patterns in an unsampled location (Sequeira, Bouchet, Yates, Mengersen, & Caley, 2018; Yates et al., 2018).

Physical surrogates such as topography, altitude, depth, broad-scale habitat types, or distance to domain boundaries can be useful for planning at regional spatial scales where detailed biological data are sparse (Bridge, Grech, & Pressey, 2015; Mellin, Bradshaw, Meekan, & Caley, 2010; Sarkar et al., 2005; Williams et al., 2009). However, care must be taken to avoid false homogeneity, where apparently similar geomorphic features support different ecological communities (Williams et al., 2009). Biological surrogates based on higher taxonomic levels (e.g., phylum) can represent biodiversity patterns more effectively than other forms of biological surrogates (Mellin et al., 2011), particularly at very large spatial scales (Last et al., 2009); however, biological data for particular taxa are often lacking in more remote areas. Predictive models emphasizing collective properties of biodiversity rather than individual entities could help overcome some of these issues associated with data deficiency and help inform effective conservation and management (Ferrier, 2002).

In recent years, there has been an increasing trend towards the designation of very large marine protected areas, or "VLMPAs," in remote locations (Singleton & Roberts, 2014). While VLMPAs have greatly increased the proportion of the oceans nominally incorporated in protected areas, questions have been raised regarding the conservation value of VLMPAs with respect to key measures such as representativeness and conservation impact (Baylis et al., 2016; Jones & De Santo, 2016; Smallhorn-West & Govan, 2018; Welch, Pressey, & Reside, 2018). A key problem with assessing the effectiveness of VLMPAs is the lack of data on biological communities at appropriate spatial and temporal scales; indeed, with the exception of shallow coastal habitats, spatial information on the distribution of biodiversity is particularly sparse in the oceans (McArthur et al., 2010; Butler, Rees, Beesley, & Bax, 2010, but see Letessier et al., 2019). Furthermore, knowledge of marine biodiversity and ongoing research tends to be concentrated in regions convenient for sampling and where funding is comparatively more available, rather than areas supporting the highest biodiversity, facing the greatest threats, or most in need of conservation (Fisher et al., 2011).

In 2012, the Australian Federal Government designated a network of marine protected areas (Australian Marine Parks; AMPs) covering 3.1 million km^2 of its exclusive economic zone (EEZ), which came into effect in 2018. Within this national framework, the North Marine Parks Network covers a large area of offshore northern Australia, including the Timor Sea, Arafura Sea, and the Gulf of Carpentaria. Located adjacent to the Coral Triangle, the region supports high biodiversity (Fisher et al., 2011; Tittensor et al., 2010), but is increasingly impacted by human activities such as offshore oil and gas extraction and fishing. The Oceanic Shoals Marine Park, named for its extensive areas of submerged carbonate terraces and banks that characterize the continental shelf of the region, is one of eight large, remote reserves in the North Marine Parks network and covers an area of 71,744 km^2 (Figure 1). Water depths across the terraces and banks (hereafter termed "shoals") range from 30 to 70 m, and are characterized by mixed sand and gravel substrata and hardground that support key habitat-forming benthic fauna, including hard corals, octocorals, and sponges, as well as a diverse assemblage of reef-associated fishes (Nichol et al., 2013; Przeslawski et al., 2011; Przeslawski et al., 2015). However, spatial heterogeneity in biodiversity patterns among and within shoals is currently poorly known, and some shoals may exhibit higher conservation value than others. Identifying the best set of shoals to protect within a large geographic area such as the Oceanic Shoals remains problematic—a challenge common to many VLMPAs.

The effectiveness of conservation planning and management at large spatial scales, particularly in data
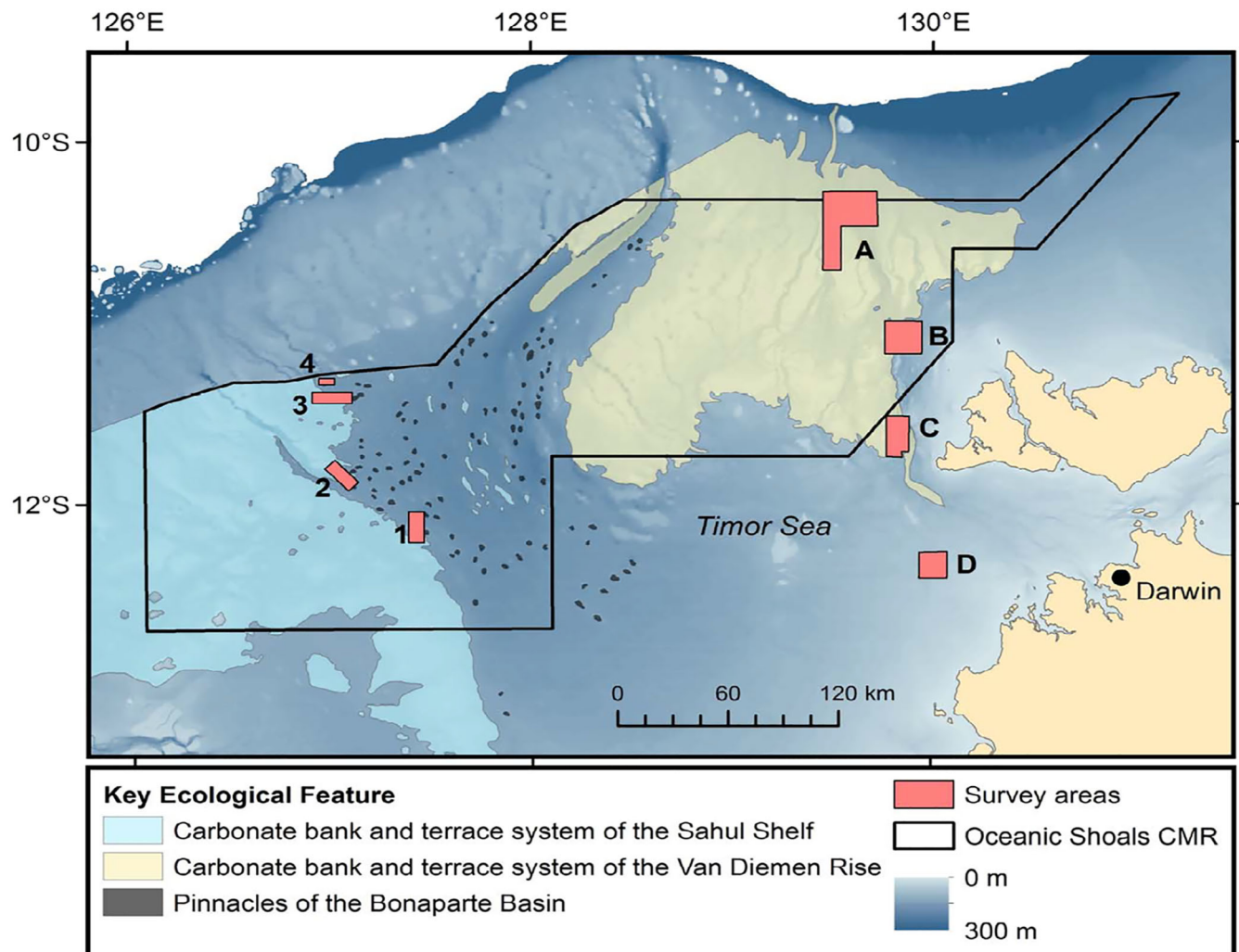
FIGURE 1 Map of study site; (a) regional map showing study region (red square); (b) close-up of the Oceanic Shoals Australian Marine Park showing location of study sites

deficient regions, could be improved considerably if predictive models from better sampled areas could be transferred to data-deficient areas to predict biodiversity patterns effectively (e.g., Sequeira et al., 2016; Sequeira et al., 2018; Yates et al., 2018). Here, we generate predictive models based on geophysical variables and ecological surveys to predict the spatial distributions of four key habitat-forming macrobenthic taxa (hard corals, soft corals, gorgonians, and sponges) in the Oceanic Shoals Marine Park. These four taxa were chosen because they are the most important habitat-forming benthos in the region, and have been shown as reliable proxies for overall biodiversity (Przeslawski et al., 2015; Przeslawski, Alvarez, Battershill, & Smith, 2014). Specifically, we investigate the extent to which (a) model predictions of biodiversity patterns across different taxa with different life histories are congruent within and among shoals; and (b) models generated in an area of extensive sampling can predict the occurrence of habitat-forming

macrobenthos in other previously surveyed regions. Predictive models that transfer well across regions could significantly assist conservation planning in regions where ecological data are scarce. Therefore, our analyses provide a novel example of the potential of such model transfers for broader application in the world's oceans.

2 | METHODS

2.1 | Data collection

Four study sites were extensively sampled in each of two regions ("West" and "East") within the Oceanic Shoals Marine Park (Figure 1) as part of three separate cruises on RV Solander. The first two cruises (SOL4934, SOL5117) sampled in the eastern region of the Oceanic Shoals Marine Park in the vicinity of the Van Diemen Rise (11.6°S, 129.7°E), and the third (SOL5650) in the

4 of 12 | WILEY— Conservation Science and Practice
A journal of the Society for Conservation Biology
Open Access

BRIDGE ET AL.

TABLE 1 Geophysical variables used in the model

Variable	Description	
Bathymetry	Bathymetry	Water depth
	Slope	Slope gradient
	Relief	Topographic relief
	Surface area	"true" surface area in relation to "planar" surface area, an indicator of surface rugosity (Jenness, 2004)
	TPI	Topographic (Benthic) Position Index (Weiss, 2001)
	Planar curvature	The curvature of the surface perpendicular to the slope direction
	Profile curvature	The curvature of the surface in the direction of slope
	Backscatter	Backscatter intensity
Backscatter	Local Moran I	An indicator of spatial autocorrelation

western region between the Malita Shelf and the Londonderry Rise (11.8°S, 127.0°E) in the Timor Sea (Figure 1). Surveys were conducted between August 27 and September 24, 2009 (SOL4934; Heap et al., 2010); July–August 2010 (SOL5117; Anderson et al., 2011); and September–October 2012 (SOL5650; Nichol et al., 2013).

Information on bathymetry and backscatter at each site was collected using multibeam sonar and used to estimate 10 geophysical variables used as explanatory variables in our predictive models (Table 1). Geophysical characteristics of the seafloor were derived from bathymetry and backscatter data collected with a Kongsberg EM3002D (300 kHz) multibeam sonar system. Raw bathymetry data were processed using the Caris™ HIPS/SIPS software. The processing involved first applying algorithms to correct for tides and vessel pitch, roll, and heave, and then using software filters and visual inspection of each swath line to remove any remaining artifacts (e.g., nadir noise and data outliers). Raw backscatter data were processed using CMST-GA MB Process v8.11.02.1, a multibeam backscatter processing toolbox co-developed by Geoscience Australia and the Centre for Marine Science and Technology at Curtin University of Technology. The backscatter processing included correction for transmission loss and ensonification area, and removal of the system-implemented model and angular dependence (Gavrilov, Siwabessy, & Parnum, 2005). Backscatter processing resulted in two separate outputs: (a) a backscatter mosaic (grid) where the equalized backscatter strengths were normalized to the backscatter strength at an incidence angle of 25°; and (b) angular backscatter response curves generated by calculating backscatter returns as a function of incidence angles (Huang, Siwabessy, Nichol, Anderson, & Brooke, 2013; Siwabessy et al., 2018). Each angular backscatter response curve

was derived from a sliding window, the size of which increased with the water depth (see Huang et al., 2013; Siwabessy et al., 2018). Consequently, the gap between adjacent points (e.g., centers of adjacent sliding windows) could be up to ~10 m along track and ~170 m across track at the deepest locations of the study area.

The remaining geophysical predictors were derived from the multibeam bathymetry data. Depth can be a powerful predictor of benthic species distributions (Gogina, Glockzin, & Zettler, 2010; McArthur et al., 2010), because it is a proxy for light availability, water temperature, and many nutrients that cannot be accurately quantified across large spatial scales. Acoustic backscatter correlates with sediment properties and substrate composition (Kloser, Bax, Ryan, Williams, & Barker, 2001; Huang et al., 2013, Huang, McArthur, et al., 2014), and can also be a useful predictor of benthic biodiversity (McArthur et al., 2010; Huang, Siwabessy, et al., 2014). Six additional geophysical variables were derived from the bathymetry grid: slope, relief, surface area (Jenness, 2004), topographic position index (Weiss, 2001), planar curvature, and profile curvature (Table 1). The morphological and terrain variables describe seabed rugosity and heterogeneity and can also be effective proxies for seabed hydrodynamic regime, which in turn can affect species distributions among patches and habitats (Holmes, Radford, & Grove, 2008; Kostylev, Erlandsson, Ming, & Williams, 2005; McArthur et al., 2010). In addition to the backscatter mosaic (grid), we also calculated local Moran's I (Anselin, 1995) to estimate the extent of spatial autocorrelation. All processed bathymetry and backscatter data, including all derived predictor variables, were gridded to a 10 × 10 m spatial resolution.

A towed sled fitted with video and digital still cameras was used to characterize benthic ecological

communities in the Oceanic Shoals CMR. The cameras fitted included a single forward-facing video camera (Watec color D250) and two still cameras, one forward-facing and one downward-facing (Sea & Sea DX2G 12 megapixel). The sled was towed at 1–1.5 knots and its position tracked using a ultra-short baseline (USBL) acoustic tracking system. Ecological data for the western region were derived from 10,785 still images collected at 10-s intervals along 52 transects each 1,500 m long, chosen using a Generalized Random Tessellation Sampling (GRTS) design with site selection weighted towards shallower areas (less than 50 m water depth) likely to support greater biodiversity (Nichol et al., 2013). Images were annotated by scoring biota and substrate beneath five random points overlaid on each image.

Ecological data from the eastern region were derived from 122 towed video transects each approximately 500 m long. The locations of transects were chosen to represent geomorphic feature types (bank, terrace, plain, and valley) interpreted from multibeam sonar mapping, with individual transects located randomly within features. Biota were recorded in real-time using the three-tiered C-BED classification scheme, which records the presence of key biota in real time for 15-s intervals every 30 s (Anderson, Van Holliday, Kloser, Reid, & Simard, 2008), and validated post-survey from georeferenced video footage combined with high-resolution photos recorded every 5 s along the video transects (Przeslawski et al., 2011). The total sample size for the eastern region was 2,703 annotated 15-s intervals. The variability in sampling design among regions was primarily due to a focus on investigating depth zonation on raised geomorphic features in the western region. In order to standardize the two different annotation techniques, four taxonomic groups (hard corals, soft corals, gorgonians, and sponges) common to both techniques were chosen, and the associated presence/absence of them recorded for each georeferenced image or video window. The vast majority of habitat-forming macrobethos in the region was contained within these four groups, resulting in a large enough sample size in each group for predictive modeling.

2.2 | Statistical analyses

The relationships between geophysical variables and the four taxonomic groups in the western region were estimated using boosted regression trees (BRTs) (Elith, Leathwick, & Hastie, 2008). BRTs are a machine-learning algorithm that incorporates multiple individual trees, and use a stage-wise approach where the largest deviation in the response variable is explained by the primary split in the tree, with remaining trees built using the residual data. This process enables identification of lower-order interactions in the predictor variables, which is not possible from other approaches that use multiple trees but average the results (Elith et al., 2008). BRTs have the additional advantage of being capable of identifying nonlinear responses to predictor variables (De'ath, 2007). All BRT models were fitted using the "gbm" package (Ridgeway, 2006) with extensions in Elith et al. (2008) in R version 3.3.3 (R Core Team, 2014).

We constructed three different models for each taxon using different parameters in order to ensure that our results were robust to variation in learning rate (i.e., that the results were not overly-influenced by the primary set of trees), tree complexity (to test for overfitting) and number of trees. Model parameters for the first set of models (tree complexity and learning rate) were assigned such that the number of trees was as close as possible to 1,500, ensuring that the models include enough trees (>1,000) to ensure stability, but also a consistent number of trees, enough to allow comparison among models (Elith et al., 2008). Tree complexity, the number of splits in each tree, was varied between three and five to prevent over-fitting while allowing the identification of interactions between variables. For all models, we used as low a learning rate as possible (0.005) to reduce the influence of the primary set of trees on the models. These initial models were then compared to two additional sets of models to examine the robustness of our results: the first set of models used an increased learning rate (0.01) to test the influence of the primary set of trees on the results, and the second set used simplified models, which utilized a reduced number of variables. The number of variables dropped varied from 2 to 6 depending on the taxon modeled, and were chosen using the "gbm.simp" function in "gbm.step."

Semivariograms were used to test for spatial autocorrelation in the occurrence dataset, which indicated an asymptote in semivariance for all four taxa at a distance of ∼70 m (Figure S1). To account for spatial autocorrelation, we added a residual autocovariate term as an additional predictor variable to explicitly account for spatial autocorrelation in the occurrence data. The residual autocovariate approach (Crase, Liedloff, & Wintle, 2012) calculates the similarity between the values of the response variable relative to neighbouring cells based on model residuals, rather than directly from the response variable. Models using residual autocovariates provide superior performance to traditional autologistic approaches because spatial autocorrelation in the response variable may be explained by autocorrelation in the predictor variables, therefore calculating the autocovariate term using model residuals allows the

TABLE 2 Summary of model results

Model	Taxon	N. tree	t. complex	L. rate	Training correlation	CV correlation	CVD	TD	DE	Training AUC	CV AUC
rac_ArbSC10	Gorgonians	1,450	3	0.005	0.494	0.402	0.248	0.353	0.30	0.925	0.908
rac_SC10	Soft corals	1,450	3	0.005	0.452	0.368	0.322	0.427	0.25	0.894	0.873
rac_Sponge10	Sponges	1850	3	0.005	0.523	0.391	0.204	0.293	0.30	0.936	0.912
rac_HC10	Hard corals	1700	3	0.005	0.752	0.668	0.124	0.286	0.57	0.964	0.957
rac_ArbSC10.Lr	Gorgonians	650	5	0.01	0.543	0.392	0.249	0.353	0.29	0.932	0.906
rac_SC10.Lr	Soft corals	550	5	0.01	0.485	0.362	0.323	0.427	0.24	0.902	0.872
rac_Sponge10.Lr	Sponges	550	5	0.01	0.55	0.381	0.206	0.293	0.30	0.94	0.912
rac_HC10.Lr	Hard corals	600	5	0.01	0.774	0.668	0.124	0.286	0.57	0.964	0.957
ArbSC.Simp	Gorgonians	1,250	5	0.005	0.541	0.4	0.248	0.353	0.30	0.933	0.907
SC.Simp	Soft corals	1,200	5	0.005	0.491	0.369	0.322	0.427	0.25	0.904	0.875
Sponge.Simp	Sponges	1,000	5	0.005	0.539	0.389	0.204	0.293	0.30	0.939	0.912
HC.Simp	Hard corals	1,200	5	0.005	0.773	0.68	0.122	0.286	0.57	0.964	0.959

explanatory variables to account for spatial autocorrelation between the occurrence data and the environmental predictors (Crase et al., 2012). Building the autocovariate term on model residuals therefore better captures the true influence of the predictor variables, and also allows delineation of spatial autocorrelation deriving from exogenous sources (e.g., geophysical variables such as depth or temperature) or endogenous sources specific to the taxon of interest (e.g., dispersal capacity; Crase et al., 2012). We calculated the autocovariate term for windows of 3 × 3 pixels surrounding each occurrence record. This window size equates to the scale at which spatial autocorrelation was evident in the occurrence data based on examination of the semivariograms (Figure S2).

Predictive performance of each model was compared using two metrics: Area Under the Receiver Operating Characteristic Curves (AUC; Fielding & Bell, 1997), and the total amount of deviance explained (DE) by each model. DE was calculated by dividing the difference between the mean total deviance (*td*) and the estimated tenfold cross-validated deviance (*cvd*) by the mean total deviance (deviance explained = 1 − *cvd*/*td*), that is, $DE = 1 - \frac{cvd}{td.}$

The proportion of the total deviance explained by each predictor variable was assessed using partial dependency plots. Model performance was evaluated using the tenfold cross-validation (see Elith et al., 2008).

For each taxon, model transferability was assessed by investigating whether habitat suitability thresholds based on BRT models for the western region could predict the presence or absence of the same taxon in the towed video surveys from the eastern region. For each BRT model, the predicted probabilities for every grid cell for each model were converted into binary (suitable/unsuitable) predictions across the entire western region. We calculated thresholds for habitat suitability using an "optimize" function that searches a parameter space via a combination of parabolic interpolation and Goldern-section search. We then use these optimized values to calculate three metrics: sensitivity (the proportion of true positives identified correctly), specificity (the proportion of true negatives), and classification rate (the proportion of positives/negatives identified correctly). Transferability was evaluated by investigating the proportion of times the habitat suitability thresholds for the models from the western region correctly predicted the location of the same taxon from the towed video surveys in the eastern region using each of the three metrics identified above. Three different metrics were used because they provide slightly different indicators of model performance; for example, where occurrence records are sparse (as in our data set and most other ecological data), predicting where something does not occur (i.e., specificity) is generally easier than predicting where it does (sensitivity), since high success in predicting absences would be expected by chance regardless of model performance. We also compared the range of values for each variable used in the model to that the range of predictors matched the range of values for both regions (Figure S2).

3 | RESULTS

The results of all three sets of models were virtually identical (Table 2, Figure S3), indicating our results were
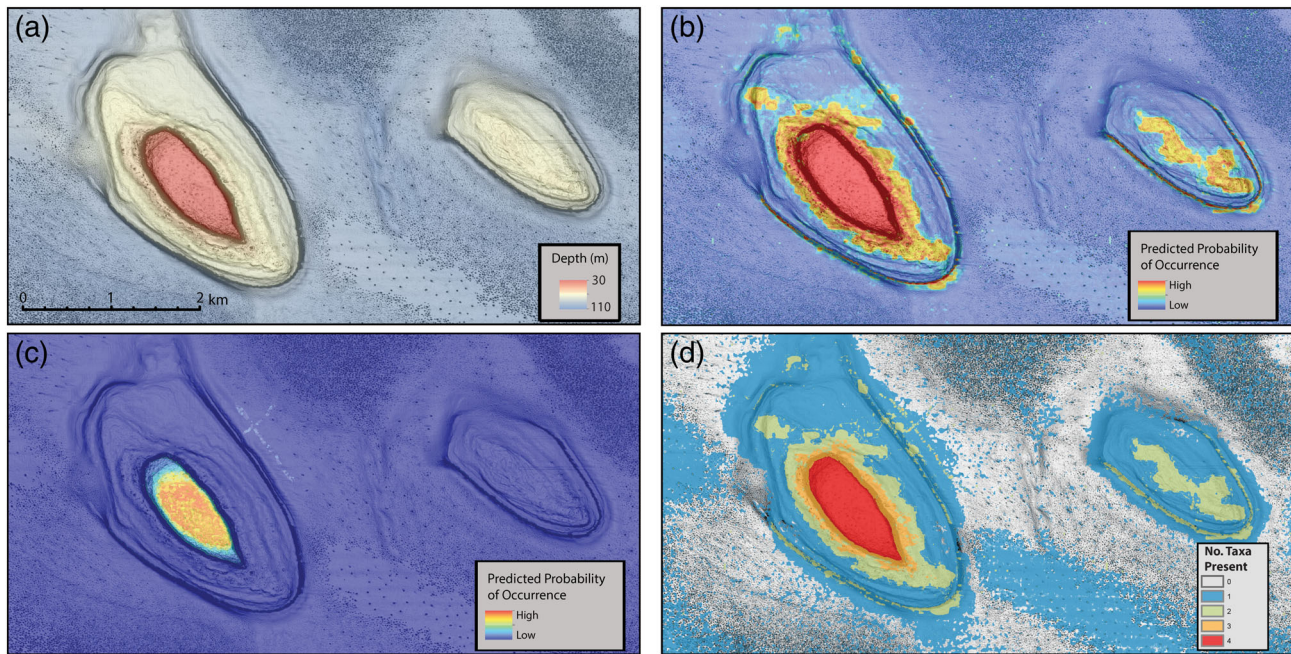
FIGURE 2 Example showing differences in predicted likelihood of occurrence between raised geomorphic features1a). Depth (a); predicted likelihood of occurrence for sponges (b); predicted likelihood of occurrence for hard corals (c); total number of taxa predicted for section of the western Oceanic Shoals AMP. Predictions cover Area 4 from regional overview (Figure 1)

robust to variability in parameters (i.e., not overly influenced by initial trees or overfit). The results for all three sets of models are provided in Table 2, but given the similarity between all sets of models, the results below refer specifically to the initial set of models. All four taxa were predicted to occur on the hard substratum on the tops of submerged shoals, but not in deeper areas with soft substrata between reefs (Figure 2). AUC values indicated good predictive performance for all models (0.87–0.94), indicating that the model was capable of discriminating between occupied and unoccupied sites (Table 2). Total deviance explained by the models varied from 24% for soft corals to 58% for hard corals (Table 2). The autocovariate term explained approximately half of the total deviance for heterotrophic taxa (gorgonians and sponges), but only ∼30% for phototrophs (hard and soft corals; Figure 3). Backscatter intensity, depth, and topographic position index were also important predictors of species' distributions.

Partial dependency plots (Figure S4) showed that the probability of occurrence for all taxa was highest at shallower depths (40–60 m) and very low at depths >80 m. All four taxa were more likely to occur on hard substrata. However, hard substrata alone did not predict suitable habitat for any of the four taxa, and models predicted low biodiversity on hard substratum on the
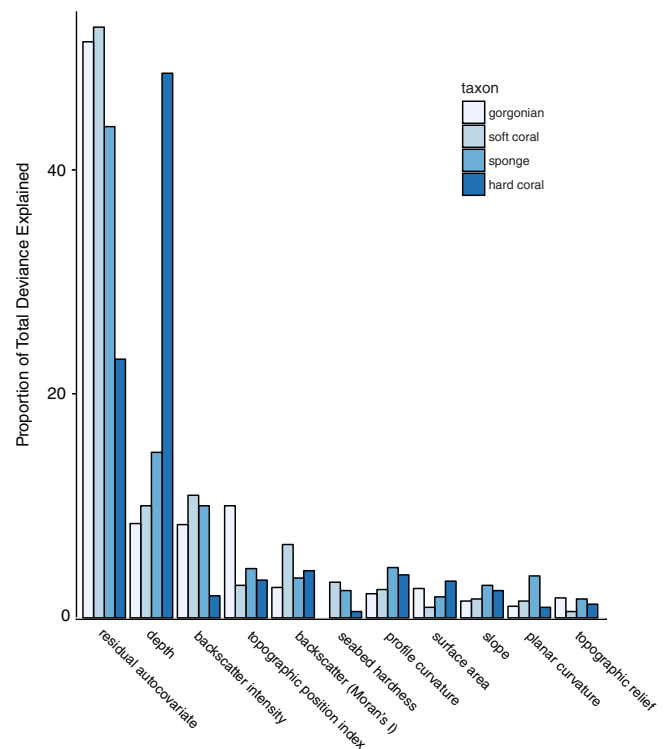

FIGURE 3 Contribution of each predictor variable to the total deviance explained by boostred regression tree models for each of the four taxa

Taxon	Suitability threshold	Sensitivity	Specificity	Classification rate
Gorgonians	0.19	0.93	0.97	0.99
Soft corals	0.13	0.93	0.95	0.95
Hard corals	0.11	0.95	0.99	0.99
Sponges	0.21	0.92	0.89	0.90

TABLE 3 Transferability of model results from western to eastern regions

Note: Four metrics were used to quantify transferability: sensitivity (the proportion of presences correctly identified); specificity (the proportion of absences correctly identified); classification rate (the proportion of both presences and absences correctly identified).

tops of deeper shoals (Figure 3). The considerable influence of the autocovariate for some taxa indicated that their distributions were highly clumped spatially. Five variables: topographic relief, slope gradient, "true" surface area, planar curvature and profile curvature of the seafloor contributed little to model predictions for any of the four taxa.

BRT models for the western region performed exceptionally well at predicting the presence and absence of the same taxon in the eastern region (Table 3). Classification rate (proportion of 1 and 0 s identified correctly) ranged from 0.90 for sponges to 0.99 for hard corals and gorgonians. For all taxa except sponges, sensitivity values were greater than specificity, indicating that the models were slightly better at predicting where a species did not occur than where it did.

4 | DISCUSSION

Our BRT models predicted well the presence or absence of a taxon outside the modeled region (i.e., they displayed high transferability). Specifically, these models transferred well for benthic communities at lower taxonomic resolution (order or higher) across 100 s of km, using a set of predictors that were largely overlapping in central tendency and variability. These transfers would therefore be best categorized as internal transfers with respect to the predictor variables used (*sensu* Sequeira et al., 2018). The success of the transfer of these models also indicates that these predictive models can support conservation planning beyond areas subjected to ecological sampling, particularly when predicting into unsampled regions with similar geophysical characteristics. This result is significant because it suggests that positive conservation outcomes could be achieved in some situations on the basis of relatively rudimentary ecological knowledge, and is particularly pertinent given the rapid implementation of large, remote MPAs and the impediments to collecting ecological data in the marine realm. Our results suggest that predictive models such as these could provide important tools to identify key habitats, or biodiversity

"hotspots," to maximize the effectiveness of large, remote reserves (see also Bouchet et al., 2017).

Despite high predictive performance across regions, our models also suggested some degree of false homogeneity among shoals: not all the shoals in the Oceanic Shoals Marine Park support equivalent biodiversity. Considerable variability in benthic composition among geomorphically-similar shoals appears to be common on the NorthWest Shelf (Heyward, 2011; Heyward, Pinceratto, & Smith, 1997; Wahab et al., 2018) and elsewhere (Bridge et al., 2011). Spatial variability, therefore, must be considered to avoid false homogeneity in representativeness and maximize the likelihood that robust predictions of biodiversity patterns are achieved in regions where considerable uncertainty exists. The extensive heterogeneity observed among shoals in the biodiversity they support indicates that the inclusion of ecological predictors of distributions may enable better predictions of biodiversity hotspots than geophysical data alone. Although raised features with hard substrata in the Oceanic Shoals MP clearly support a greater abundance and diversity of macrobenthos than adjacent areas of soft substrata, not all raised features, or parts thereof, were equally likely to support abundant macrobenthos. These modeled results are consistent with observations for sponges based on specimen collections (Przeslawski et al., 2014, 2015) with high variability in community composition among and between shoals. Therefore, fine-scale variability within and among apparently similar geomorphic features should also be considered to maximize the conservation benefits of reserves designs where ecological data are sparse.

Depth and variables associated with substrate type best predicted the occurrence of habitat-forming benthic macrofauna, indicating their preference for the tops of seabed features where hard substrata are present (Figure 3). The importance of depth for predicting biodiversity patterns in the Oceanic Shoals Marine Parksupports their use in Australia's national marine bioregionalisation (Last et al., 2009). Unsurprisingly, depth was a particularly good predictor of phototrophic taxa, where banks with tops shallower than 45 m supported

greater biodiversity than deeper banks. This result is consistent with Williams et al. (2009), who also reported that depth generated considerable heterogeneity in ecological communities among similar geomorphic features over a much larger depth range (thousands of meters) than explored here.

Spatial autocorrelation is common in ecological data, and may arise from both exogenous environmental factors such as rainfall, or depth, and endogenous ones, such as species-specific factors including dispersal limitation (Crase et al., 2012; Keitt, Bjørnstad, Dixon, & Citron-Pousty, 2002; Legrende, 1993; Legendre et al., 2002). The importance of the spatial autocovariate indicates that the clumped distributions of taxa observed here were strongly influenced by factors other than the broad-scale environmental predictors used. Potential causes of these patterns could include ecological factors (e.g., settlement processes) or additional environmental factors not considered here (Keitt et al., 2002), such as disturbance history, oceanography, and turbidity. For example, gorgonians rely on currents for food, and commonly occur in areas of high current (Fabricius & Alderslade, 2001). Most gorgonians are also brooders, where planulae typically settle within a few meters of the parent colony. The patchy distribution of gorgonians in the Oceanic Shoals MP may therefore be due to a variety of ecological and environmental factors that cannot be identified from broad-scale geophysical data, including variability in current flows both among and within shoals and/or their reproductive ecology. Accounting for this spatial autocorrelation in predictive models can help predict patchy species occurrences even if the underlying causes are unknown.

Although all four focal taxa showed similar habitat preferences, our models investigated patterns of biodiversity at a low taxonomic resolution of order, or above, to ensure reasonable sample sizes and potentially improve their utility as biological surrogates. Many of these taxa might exhibit greater spatial variability if greater taxonomic resolution was applied. Spatial distribution patterns for other taxa are poorly known, and our "coarse-filter" approach provides no information on the suitability of surrogates for planning specifically for species that may be more in need of conservation action. Identification of particular species or habitats of conservation concern (e.g., spawning aggregation sites) should therefore be an additional consideration when designing any marine reserve network (Hamilton, Potuku, & Montambault, 2011; Weeks et al., 2014).

Our results provide evidence that, at low taxonomic resolution, protecting raised features with hard substrate will benefit a wide range of taxa in the Oceanic Shoals MP. Importantly, models based on one region successfully predicted species presence in another, suggesting predictive models have considerable value for conservation planning in data-poor regions over large spatial scales. However, even at low taxonomic resolution heterogeneity among similar geomorphic habitats within a region can be high. Consequently, additional data, including ongoing monitoring, may be required to address more specific and finer-scale conservation and management goals. Nonetheless, our results corroborate previous studies that demonstrate predictive models can provide valuable information on the distribution of biodiversity across large geographic areas. Importantly, our models of the distributions of benthic taxa showed greater capacity for transferability than previous studies based on fishes (Sequieira et al., 2016). Given that biodiversity data are often scarce, particularly in biodiversity-rich areas most in need of conservation, our findings provide further support for the notion that conservation benefits can result from marine reserves designed on the basis of relatively simple geophysical data with careful consideration of the potential for false homogeneity and the transfer of predictive models among locations.

CONFLICT OF INTEREST

The authors declare no conflict of interest

AUTHOR CONTRIBUTIONS

Tom C. L. Bridge and M. Julian Caley conceived the study. Zhi Huang, Rachel Przeslawski, Maggie Tran, Justy Siwabessy, Kim Picard, and Scott L. Nichol collected and processed the raw data, Tom C. L. Bridge, April E. Reside, and Murray Logan conducted the analysis. Tom C. L. Bridge wrote the paper, with input from all other authors. All authors were involved in providing feedback on the analysis and revising the manuscript. All authors approved the final submission.

DATA AVAILABILITY STATEMENT

Bathymetry data for the Oceanic Shoals Marine Park is publically available from Geoscience Australia at https://data.gov.au/data/dataset/f6311fa7-6ddd-44dc-bbf3-6e7560bbfe21. Seabed imagery is available from the Australian Marine Video and Imagery Collection housed on the National Computational Infrastructure https://cmi.ga.gov.au/australian_marine_video_and_imagery_collection. Code to reproduce the analysis is available from the corresponding author upon request.

ORCID

Tom C. L. Bridge https://orcid.org/0000-0003-3951-284X

REFERENCES

Anderson, J. T., Van Holliday, D., Kloser, R., Reid, D. G., & Simard, Y. (2008). Acoustic seabed classification: Current practice and future directions. *ICES Journal of Marine Science*, 65 (6), 1004–1011.

Anderson, T. J., Nichol, S., Radke, L., Heap, A. D., Battershill, C., Hughes, M., ... Daniell, J. (2011). Seabed environments of the eastern Joseph Bonaparte gulf, northern Australia: GA0325/Sol5177—Post-survey report. Geoscience Australia, Record 2011/08, 59pp.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115.

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., ... Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9(1), 58–64.

Bouchet, P. J., Meeuwig, J. J., Huang, Z., Letessier, T. B., Nichol, S. L., Caley, M. J., & Watson, R. A. (2017). Continental-scale hotspots of pelagic fish abundance inferred from commercial catch records. *Global Ecology and Biogeography*, 26, 1098–1111.

Bridge, T. C., Done, T. J., Friedman, A., Beaman, R. J., Williams, S. B., Pizarro, O., & Webster, J. M. (2011). Variability in mesophotic coral reef communities along the great barrier reef, Australia. *Marine Ecology Progress Series*, 428, 63–75.

Bridge, T. C. L., Grech, A. M., & Pressey, R. L. (2015). Factors influencing incidental representation of previously unknown features in marine protected areas. *Conservation Biology*, 30, 154–165.

Butler, A. J., Rees, T., Beesley, P., & Bax, N. J. (2010). Marine biodiversity in the Australian region. *PLoS One*, 5(8), e11831. https://doi.org/10.1371/journal.pone0011831

Crase, B., Liedloff, A. C., & Wintle, B. A. (2012). A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, 35, 879–888.

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, 88, 243–251.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813.

Fabricius, K. K., & Alderslade, P. P. (2001). *Soft corals and sea fans: A comprehensive guide to the tropical shallow water genera of the central-West Pacific, the Indian Ocean and the Red Sea*. Townsville, Australia: Australian Institute of Marine Science (AIMS).

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, 51, 331–363.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38–49.

Fisher, R., Radford, T., Knowlton, N., Brainard, R. E., Michaelis, F. B., & Caley, M. J. (2011). Global mismatch between research effort and conservation needs of tropical coral reefs. *Conservation Letters*, 4, 64–72.

Gavrilov, A. N., Siwabessy, P. J. W., and Parnum, I. M. (2005) Multibeam echo sounder backscatter analysis: theory review, methods and application to Sydney Harbour swath data. CRC Milestone Report CA3.03. CMST report 2005–03 for the Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management. Centre for Marine Science and Technology, Curtin University of Technology, Perth, Australia.

Gogina, M., Glockzin, M., & Zettler, M. L. (2010). Distribution of benthic macrofaunal communities in the western Baltic Sea with regard to nearbottom environmental parameters. 2. Modelling and prediction. *Journal of Marine Systems*, 80, 57–70.

Grantham, H. S., Pressey, R. L., Wells, J. A., & Beattie, A. J. (2010). Effectiveness of biodiversity surrogates for conservation planning: Different measures of effectiveness generate a kaleidoscope of variation. *PLoS ONE*, 5(7), e11430. https://doi:10.1371/journal.pone.0011430.

Hamilton, R. J., Potuku, T., & Montambault, J. R. (2011). Community-based conservation results in the recovery of reef fish spawning aggregations in the coral triangle. *Biological Conservation*, 144(6), 1850–1858.

Heap, A.D., Przeslawski, R., Radke, L., Trafford, J., Battershill, C. & Shipboard Party. (2010). Seabed Environments of the Eastern Joseph Bonaparte Gulf, Northern Australia. SOL4934—Post-survey Report. Geoscience Australia, Record 2010/09, 78pp.

Heyward, A. A., Pinceratto, E. E., & Smith, L. L. (1997). *Big Bank Shoals of the timor sea: An environmental resource atlas*. Chicago, USA: Australian Institute of Marine Science & BHP Petroleum.

Heyward, A., Jones, R., Meeuwig J., Burns K., Radford B., Colquhoun J, ... Stowar, M. (2011) Monitoring study S5 banks & shoals, Montara 2011 Offshore Banks Assessment Survey. Report for PTTEP Australasia (Ashmore Cartier) Pty Ltd. Australian Institute of Marine Science, Townsville, Australia. Available from https://www.environment.gov.au/system/files/pages/bcefac9b-ebc5-4013-9c88-a356280c202c/files/2011-offshore-banks-assessment-survey.pdf.

Holmes, K. W., VanNeil, K. P., Radford, B., Kendrick, G. A., & Grove, S. L. (2008). Modelling distribution of marine benthos from hydroacoustics and underwater video. *Continental Shelf Research*, 28, 1800–1810.

Huang, Z., McArthur, M., Przeslawski, R., Siwabessy, P. J. W., Nichol, S., Anderson, T., & Brooke, B. (2014). Predicting mapping of soft bottom benthic biodiversity using a surrogacy approach. *Marine and Freshwater Research*, 65, 409–424.

Huang, Z., Siwabessy, J., Nichol, S., Anderson, T., & Brooke, B. (2013). Predictive mapping of seabed cover types using angular response curves of multibeam backscatter data: Testing

different feature analysis approaches. *Continental Shelf Research*, *61*, 12–22.

Huang, Z., Siwabessy, J., Nichol, S. L., & Brooke, B. P. (2014). Predictive mapping of seabed substrata using high-resolution multibeam sonar data: A case study from a shelf with complex geomorphology. *Marine Geology*, *357*, 37–52.

Jenness, J. S. (2004). Calculating landscape surface area from digital elevation models. *Wildlife Society Bulletin*, *32*, 829–839.

Jones, P. J. S., & De Santo, E. M. (2016). Is the race for remote, very large marine protected areas (VLMPAs) taking us down the wrong track? *Marine Policy*, *73*, 231–234.

Keitt, T. H., Bjørnstad, O. N., Dixon, P. M., & Citron-Pousty, S. (2002). Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*, *25*, 616–625.

Kloser, R. J., Bax, N., Ryan, J. T., Williams, A., & Barker, B. A. (2001). Remote sensing of seabed types in the Australian south east fishery; development and application of normal incident acoustic techniques and associated 'ground truthing'. *Marine and Freshwater Research*, *52*, 475–489.

Kostylev, V. E., Erlandsson, J., Ming, M. Y., & Williams, G. A. (2005). The relative importance of habitat complexity and surface area in assessing biodiversity: Fractal application on rocky shores. *Ecological Complexity*, *2*, 272–286.

Last, P. R., Lyne, V. D., Williams, A., Davies, C. R., Butler, A. J., & Yearsley, G. K. (2009). A hierarchical framework for classifying seabed biodiversity with application to planning and managing Australia's marine biological resources. *Biological Conservation*, *143*, 1675–1686.

Legendre, P., Dale, M. R., Fortin, M. J., Gurevitch, J., Hohn, M., & Myers, D. (2002). The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, *25*, 601–615.

Legrende, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology*, *74*, 1659–1673.

Letessier, T. B., Mouillot, D., Bouchet, P. J., Vigliola, L., Fernandes, M. C., Thompson, C., ... Meeuwig, J. J. (2019). Remote reefs and seamounts are the last refuges for marine predators across the Indo-Pacific. *PLoS Biology*, *17*, e3000366.

Lombard, A. T., Cowling, R. M., Pressey, R. L., & Rebelo, A. G. (2003). Effectiveness of land classes as surrogates for species in conservation planning for the cape floristic region. *Biological Conservation*, *112*, 45–62.

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*, 243–253.

McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer, V. L., Nichol, S., ... Radke, L. C. (2010). On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuarine, Coastal and Shelf Science*, *88*(1), 21–32.

Mellin, C., Bradshaw, C. J. A., Meekan, M. G., & Caley, M. J. (2010). Environmental and spatial predictors of species richness and abundance in coral reef fishes. *Global Ecology and Biogeography*, *19*, 212–222.

Mellin, C., Delean, S., Caley, J., Edgar, G., Meekan, M., Pitcher, R., ... & Bradshaw, C., (2011). Effectiveness of biological surrogates for predicting patterns of marine biodiversity: A global meta-analysis. *PLoS One*, *6*(6), e20141. https://doi.org/10.1371/journal.pone.0020141.

Mellin, C., Mengersen, K., Bradshaw, C. J. A., & Caley, M. J. (2014). Generalizing the use of geographical weights in biodiversity modelling. *Global Ecology and Biogeography*, *23*, 1314–1323.

Nichol, S., Howard, F., Kool, J., Stowar, M., Bouchet, P., Radke, L., ... Colquhoun, J. (2013). Oceanic Shoals Commonwealth Marine Reserve (Timor Sea) Biodiversity Survey: GA0339/SOL5650—Post Survey Report. Record 2013/38. Geoscience Australia: Canberra. Available from http://www.ga.gov.au/corporate_data/76658/Rec2013_038.pdf

Przeslawski, R., Alvarez, B., Battershill, C., & Smith, T. (2014). Sponge biodiversity and ecology of the Van Diemen rise and eastern Joseph Bonaparte gulf, northern Australia. *Hydrobiologia*, *730*, 1–16.

Przeslawski, R., Alvarez, B., Kool, J., Bridge, T., Caley, M. J., & Nichol, S. (2015). Implications of sponge biodiversity patterns for the management of a marine reserve in northern Australia. *PLoS One*, *10*(11), e0141813. https://doi.org/10.1371/journal.pone.0141813

Przeslawski, R., Daniell, J., Anderson, T., Vaughn Barrie, J., Heap, A., Hughes, M., ... Siwabessy, J. (2011) Seabed habitats and hazards of the Joseph Bonaparte gulf and Timor Sea, Northern Australia. Geoscience Australia, Record 2011/40, 69pp. (Available from http://www.ga.gov.au/corporate_data/72805/Rec2011_040.pdf)

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing http://www.R-project.org/

Ridgeway, G. (2006) Generalized boosted regression models. Documentation on the R Package 'gbm', version 1· 5, 7.

Sarkar, S., Justus, J., Fuller, T., Kelley, C., Garson, J., & Mayfield, M. (2005). Effectiveness of environmental surrogates for the selection of conservation area networks. *Conservation Biology*, *19*, 815–825.

Sequeira, A. M. M., Bouchet, P. J., Yates, K. L., Mengersen, K., & Caley, M. J. (2018). Transferring biodiversity models for conservation: Opportunities and challenges. *Methods in Ecology and Evolution*, *9*, 1250–1264.

Sequeira, A. M., Mellin, C., Lozano-Montes, H. M., Vanderklift, M. A., Babcock, R. C., Haywood, M. D., & Caley, M. J. (2016). Transferability of predictive models of coral reef fish species richness. *Journal of Applied Ecology*, *53*, 64–72.

Singleton, R. L., & Roberts, C. M. (2014). The contribution of very large marine protected areas to marine conservation: Giant leaps or smoke and mirrors? *Marine Pollution Bulletin*, *87*, 7–10.

Siwabessy, P. J. W., Tran, M., Picard, K., Brooke, B. P., Huang, Z., Smit, N., ... Atkinson, I. (2018). Modelling the distribution of hard seabed using calibrated multibeam acoustic backscatter data in a tropical, macrotidal embayment: Darwin harbour, Australia. *Marine Geophysical Research*, *39*, 249–269. https://doi.org/10.1007/s11001-017-9314-7

Smallhorn-West, P., & Govan, H. (2018). Towards reducing misrepresentation of national achievements in marine protected area targets. *Marine Policy*, *97*, 127–129.

Sutcliffe, P. R., Mellin, C., Pitcher, C. R., Possingham, H. P., & Caley, M. J. (2014). Regional-scale patterns and predictors of species richness and abundance across twelve major tropical inter-reef taxa. *Ecography*, *37*, 162–171.

12 of 12 | WILEY— Conservation Science and Practice
A Journal of the Society for Conservation Biology • Open Access

BRIDGE ET AL.

Sutcliffe, P. R., Pitcher, C. R., Caley, M. J., & Possinghan, H. P. (2012). Biological surrogacy in tropical seabed assemblages fails. *Ecological Applications*, *22*, 1762–1771.

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, *466*, 1098–1101.

Wahab, M. A. A., Radford, B., Cappo, M., Colquhoun, J., Stowar, M., Depczynski, M., ... Heyward, A. (2018). Biodiversity and spatial patterns of benthic habitat and associated demersal fish communities at two tropical submerged reef ecosystems. *Coral Reefs*, *37*(2), 327–343.

Weeks, R., Aliño, P. M., Atkinson, S., Beldia, P., Binson, A., Campos, W. L., ... Jumin, R. (2014). Developing marine protected area networks in the coral triangle: Good practices for expanding the coral triangle marine protected area system. *Coastal Management*, *42*(2), 183–205.

Weiss, A. D. (2001). *Topographic Position and Landforms Analysis*. ESRI International User Conference, San Diego, CA.

Welch, H., Pressey, R. L., & Reside, A. E. (2018). Using temporally explicit habitat suitability models to assess threats to mobile species and evaluate the effectiveness of marine protected areas. *Journal for Nature Conservation*, *41*, 106–115.

Williams, A., Bax, N., Kloser, N., Althaus, F., Barker, B., & Keith, G. (2009). Australia's deep-water reserve network: Implications of false homogeneity for classifying abiotic surrogates of biodiversity. *ICES Journal of Marine Science: Journal du Conseil*, *66*, 214–224.

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., ... Sequeira, A. M. M. (2018). Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, *33*, 790–802.

Yates, K. L., Mellin, C., Caley, M. J., Radford, B. T., & Meeuwig, J. J. (2016). Models of marine fish biodiversity: Assessing predictors from three habitat classification schemes. *PLoS One*, *11*, e0155634.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.