BIODIVERSITY RESEARCH

Geographic range determinants of two commercially important marine molluscs

Camille Mellin[1,2]*, Bayden D. Russell[3], Sean D. Connell[3], Barry W. Brook[2] and Damien A. Fordham[2]

[1]*Australian Institute of Marine Science, PMB No.3, Townsville MC, Townsville, Qld 4810, Australia,* [2]*The Environment Institute and School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia,* [3]*Southern Seas Ecology Laboratories, School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia*

ABSTRACT

Aim We modelled the spatial abundance patterns of two abalone species (*Haliotis rubra* Donovan 1808 and *H. laevigata* Leach 1814) inhabiting inshore rocky reefs to better understand the importance of current sea surface temperature (SST) (among other predictors) and, ultimately, the effect of future climate change, on marine molluscs.

Location Southern Australia.

Methods We used an ensemble species distribution modelling approach that combined likelihood-based generalized linear models and boosted regression trees. For each modelling technique, a two-step procedure was used to predict: (1) the current probability of presence, followed by (2) current abundance conditional on presence. The resulting models were validated using an independent, spatially explicit dataset of abalone abundance patterns in Victoria.

Results For both species, the presence of reef was the main driver of abalone occurrence, while SST was the main driver of spatial abundance patterns. Predictive maps at *c.* 1-km resolution showed maximal abundance on shallow coastal reefs characterized by mild winter SSTs for both species.

Main conclusions Sea surface temperature was a major driver of abundance patterns for both abalone species, and the resulting ensemble models were used to build fine-resolution predictive range maps (*c.* 1 km) that incorporate measures of habitat suitability and quality in support of resource management. By integrating this output with structured spatial population models, a more robust understanding of the potential impacts of threatening human processes such as climate change can be established.

Keywords

Abalone, boosted regression trees, generalized linear models, *Haliotis*, sea surface temperature, species distribution modelling.

*Correspondence: Camille Mellin, Australian Institute of Marine Science, PMB No.3, Townsville MC, Townsville, Qld 4810, Australia.
E-mail: c.mellin@aims.gov.au

INTRODUCTION

Species distribution models (SDMs) provide a useful tool for establishing key climate and environmental predictors of species' range. By capturing the implicit processes that limit the present-day distribution of a species, SDMs can be used to predict its probability of occurrence and abundance onto spatial maps (Guisan & Thuiller, 2005; Hirzel & Le Lay, 2008). Their use in quantifying spatial patterns of habitat suitability and geographic range size, and how these may change in response to human-related disturbances such as habitat loss and, more recently, climate change, has resulted in SDMs being commonly used to assess the extinction risk of species and

theirs habitats and inform reserve design (e.g., Araujo & Williams, 2000). Indeed, extinction risk is strongly related to geographic range size and how it changes through time (Gaston & Fuller, 2009).

Although SDMs have received increasing attention over the last decade, with the development of both correlative (see review by Guisan & Zimmermann, 2000; Elith *et al.*, 2006) and mechanistic models (i.e. process-based; Kearney & Porter, 2009), their application to date has been focused on the terrestrial domain (see review by Robinson *et al.*, 2010; Mellin *et al.*, 2010). This is partly because, until recently, there has been a paucity of high-resolution and broadscale environmental data in the marine realm (e.g. Tyberghein *et al.*, 2011).

(left margin) **Diversity and Distributions** · A Journal of Conservation Biogeography

However, recent technological advances, such as multibeam echo-sounding, backscatter imaging (e.g. McArthur *et al.*, 2010), matched with developments in remote sensing (e.g. Mellin *et al.*, 2009), can now bridge the gap between model requirements and data availability, provided that high-resolution environmental datasets are assembled in a comprehensive and uniform way (Tyberghein *et al.*, 2011).

The relative importance of certain biotic factors (e.g. dispersal and feeding) that need to be considered when applying SDMs varies between organisms; for example, ontogenetic shifts or feeding interactions can be major drivers of the distribution of marine organisms, yet they remain poorly understood for many species (e.g. Robinson *et al.*, 2010), and where established, their role in dictating range and abundance is quite challenging to account for in SDM applications. Considering that the difficulty in constructing SDMs varies among species – mainly depending on their life-history traits – marine molluscs, such as blacklip and greenlip abalone (*Haliotis rubra* and *H. laevigata*, respectively), may provide ideal candidate species to pioneer advances in marine SDMs owing to (1) a feeding behaviour focused on a non-limiting resource (i.e. drifting algae; Shepherd, 1973), (2) a restricted larval dispersal (i.e. 100s m; Prince *et al.*, 1987) and (3) a benthic adult phase with limited mobility (e.g. Shepherd, 1973). Collectively, these biotic traits also suggest a high sensitivity to changes in environmental conditions (Berg *et al.*, 2010; Statzner & Beche, 2010; Vandewalle *et al.*, 2010). The rare, already established, examples of SDMs developed for marine molluscs (Sanderson *et al.*, 2005; Inglis *et al.*, 2006; Gutierrez *et al.*, 2008) used a single modelling technique, without any independent validation of the subsequent models. Owing to a limited number of marine invertebrate case studies, additional testing and validation are required to determine the favourability of using SDMs to predict spatial patterns of marine mollusc abundance. Validation using independent data is a rigorous method for determining model performance (Botkin *et al.*, 2007).

Haliotis rubra and *H. laevigata* are the two commercially harvested abalone in southern Australia. This fishery has been regulated since 1964, with an annual total allowable commercial catch (TACC) in place since *c.* 1990, and was worth $30 million in 2009 (EconSearch 2010). Despite this, catch per unit effort has decreased substantially over the last decade in some of the more heavily fished management zones (Mayfield *et al.*, 2008, 2009; Chick *et al.*, 2009), and fisheries managers are attempting to identify new methods to predict population distribution and structure. Attempts, to date, have been compromised, as a result of a paucity of fine-resolution GIS layers, describing key environmental variables that limit abundance and the uncertainty regarding global warming and forecast change in the sea surface temperature (SST), which drives most biological and ecological processes in benthic organisms (e.g., Poloczanska *et al.*, 2008; Ling *et al.*, 2009). The two abalone species have overlapping distributions, and their reproductive biology, growth and population dynamics are well established (e.g., Shepherd, 1973; Temby

et al., 2007; Saunders *et al.*, 2009a). However, across their distributional range, population densities of each species vary according to a range of environmental variables such as temperature, habitat type and topography, and food availability (Shepherd, 1973; Saunders *et al.*, 2009b). Yet, until recently, high-resolution spatial layers describing the primary environmental drivers of abalone range and abundance were unavailable, limiting our ability to predict *H. rubra* and *H. laevigata* population densities at fine scales (e.g. ≤ 1 km^2) across either species' range.

In this study, we (1) identify key environmental drivers that explain patterns of presence and abundance across the range of two marine molluscs, (2) develop an ensemble modelling approach to explore uncertainty in different modelling methods and (3) construct high-resolution predictive maps of abundance that can be used in population models.

METHODS

Study area

The study area covered 7579 km of coastline of southern Australia, between 129–151° E and 31–40° S (Fig. 1). It is characterized by open coasts but also includes two gulfs and several large bays. Along the open coast, rocky reef often extends from the intertidal zone to 20–30 m depth, with offshore reefs of varying depth to > 50 m. In contrast, the gulfs and bays are dominated by seagrass and sand, generally at < 30 m depth, with a smaller number of rocky-reef outcrops. Reef geology varies throughout the study area, but most can be characterized according to low (limestone reef) and high (granite reef) complexity. Reefs in this region support diverse algal assemblages dominated by canopy-forming algae in the orders Fucales (fucoid algae) and Laminariales (kelps), which provide habitat structure and food for associated invertebrates and fish.

Survey methods

The abundance of *Haliotis rubra* (blacklip abalone) and *H. laevigata* (greenlip abalone) was surveyed by SCUBA (Self Contained Underwater Breathing Apparatus) diving between 5 and 30 m depth in both South Australia and Victoria. In South Australia, two types of surveys were done: annual surveys (1980–2009) that specifically targeted commercially harvested reefs and semi-annual surveys (2004–2009) of randomly chosen reefs across the entire state. Targeted surveys of harvested reefs consisted of timed swims, where a SCUBA diver swam along a predetermined compass bearing for 10 min while counting all abalone in a 1-m-wide belt ($n = 2$ timed swims per site). Non-targeted surveys were done on randomly chosen reefs, with abalone counted within either 100 × 1 m ($n = 2$ per site), 30 × 1 m ($n = 6$ per site), or 50 × 1-m ($n = 1$ per site) belt transects. In Victoria, surveys of abalone abundance were done using similar methodology to the non-targeted surveys in South Australia but at pre-selected reefs
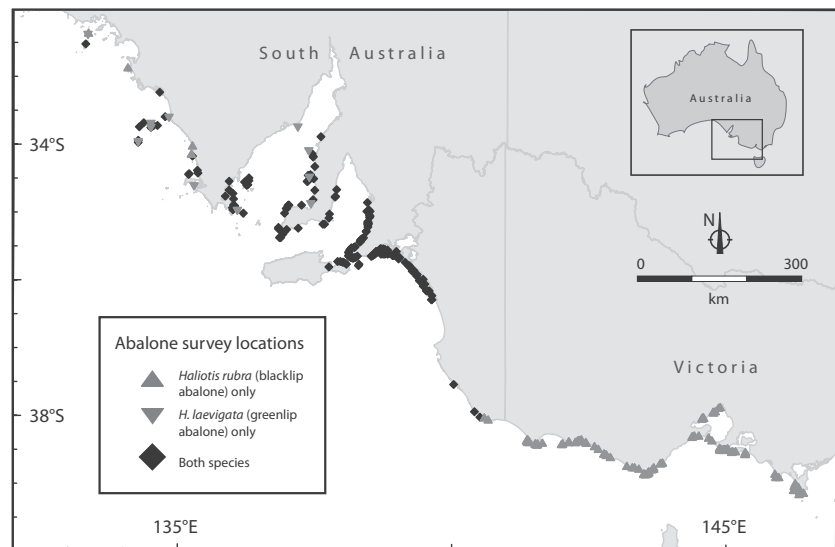
Figure 1 Abalone survey locations (projection: Geocentric Datum of Australia 1994).

that included both harvested and non-harvested reefs. Surveys were done semi-annually (1992–2009) using 30×1-m belt transects ($n = 6$–8 per site). We calculated the mean density of abalone (individuals m^{-2}) for each site by dividing the total abundance per transect by the length (m) of the transect. Densities (m^{-2}) were then multiplied by 100 to get an estimate of abundance over a 100-m^2 survey area, thereby allowing the use of a Poisson distribution for modelling non-zero count data (see Modelling). The 100-m^2 area was chosen because it was the most commonly surveyed in our data, therefore minimizing the bias that can arise when scaling densities over different areas.

Climate and environmental predictors

We extracted gridded SST data from the advanced very high-resolution radiometer (AVHRR) NASA/NOAA Pathfinder product version 5.0 (http://podaac.jpl.nasa.gov/DATA_CAT-ALOG/avhrrinfo.html), which uses an algorithm, developed by Walton *et al.* (1998), that is effective at reproducing *in situ* SST measurements (Marullo *et al.*, 2007). Pathfinder v5 data are available daily from 1985 to present at a 4.6-km grid resolution, for day and night passes separately. More specifically, we used Pathfinder v5 mean monthly day and night SST data from 1985 to 2004 to calculate a 20-year monthly day/night average and standard deviation (interannual) for March, August and annual SST at 40155 observation sites. Thin-plate spline surface-smoothing techniques were used to downscale the 4.6-km resolution Pathfinder v5 data to a 0.01°latitude/longitude grid (Wahba, 1990). The degree of smoothness of the fitted function was determined by minimizing a measure of predictive error of the fitted surface given by the generalized cross-validation (Hutchinson, 2006). Out-of-source sampling, based on a subset of the Pathfinder v5 measurements retained for validation, was used to evaluate model fit (greater than $\pm 0.5\ ^{\circ}$C). Thin-plate smoothing spline-fitting techniques are frequently used to interpolate

monthly mean climate data, with good success (e.g. Hutchinson, 1995). Though, their accuracy and precision can be influenced somewhat by the spatial distribution and density of the measurement points. However, this was not an issue in our study because all measurement points (within the study extent) were located no further than 5 km from its nearest neighbouring point.

Data on the probability of inshore reef occurrences in South Australia were constructed using bathymetric data and artificial neural networks at a 250-m resolution (Watts *et al.*, 2011). This was used to discriminate reef from non-reef subtidal habitats. A threshold of 0.8 was applied to the probability of reef occurrence, and the subsequent binary layer (1 or high: probability ≥ 0.8; 0 or low: probability < 0.8) was aggregated to a 0.01°latitude/longitude resolution. At this coarser resolution, a cell was classed as reef if at least 25% of fine-resolution cells falling within the 0.01×0.01°cell boundary had a high probability of being reef (i.e. at least four cells or 25% of cells were equal to 1). The 25% rule-of-thumb was chosen because it maximized the precision of the resulting reef layer, i.e. 83% of all known reefs were correctly classified.

To calculate fishing effort, data on the commercial fishing catch of abalone were sourced from the government organisations which regulate the state fisheries (Department of Primary Industries and Resources, South Australia and Department of Primary Industries, Victoria). The commercial harvest of abalone was collated within harvest zones in both South Australia (39 zones) and Victoria (193 zones) to give spatially explicit annual harvest (tonnes) for both *H. rubra* and *H. laevigata* from 1995 to 2008, a time period overlapping with the survey period and climate predictor information. Harvest zones are spatial management units used to allow effective control of harvest rates. While the sizes of these harvest zones vary in size from *c.* 20–200 km of coast (most zones are *c.* 40 km of coast, i.e. the fine-scale 'map codes' in Fig. 1 of Nobles *et al.* (2004)), they allow spatial analysis of the harvest. Measures of fishing intensity, calculated for each harvest zone,

were total annual harvest, mean annual harvest (*Harvest*) and interannual standard deviation, and the percentage of years each harvest zone was fished.

Mean depth (*Depth*; in m) was obtained from a *c.* 250-m resolution multibeam bathymetric dataset provided by Geoscience Australia (Webster & Petkovic, 2005). The geographic coordinates of boat ramps used by both recreational and commercial abalone fishing vessels were collated from the Departments of Fisheries of South Australia & Victoria, and the Euclidean distance to the nearest boat ramp (*Distance*; in degrees) was calculated for every grid cell in ArcGIS 9.2. As no data are available for recreational fishing pressure on abalone stocks, *Distance* was used as a proxy for this recreational fishing harvest.

All climate and environmental predictors were collated across a common 0.01° resolution grid extending over the continental shelf off South Australia and Victoria (see Fig. S1 in Supporting Information). Because *H. rubra* and *H. laevigata* occur within the 0–30 m-depth range (Shepherd, 1973), grid cells outside this range were masked out, resulting in a final grid of 40,229 cells (Table 1).

Modelling

Overview of the modelling framework

The abundance of each species was modelled using an ensemble approach including two common, easy to implement and powerful SDM techniques (Elith *et al.*, 2006), namely generalized linear models (GLM; e.g. Guisan *et al.*, 2002; Generalized linear models) and boosted regression trees (BRT; Elith *et al.*, 2008; Boosted regression trees). We used a multimodel ensemble approach (e.g. Araujo & New, 2007; Multimodel ensemble predictions) to account for the inherent uncertainty and subsequent error of each technique, both in terms of variation between different model assumptions and algorithms, and in terms of variation between the resulting models and fitted parameters (Barry & Elith, 2006). Within each technique, modelling abundance was done in two steps (Two-step modelling of zero-inflated data), i.e. modelling the probability of presence first followed by the abundance conditional on presence. Models were validated using an independent dataset (for *H. rubra* only because of restricted data availability; Out-of-sample validation), as well as standard tenfold cross-validation, and used to derive multimodel ensemble predictions across the study area (Multimodel ensemble predictions).

Two-step modelling of zero-inflated data

A two-step modelling approach was used to accommodate for the zero-inflated nature of abalone abundance records, and because the predictors of presence can be different to those of abundance (Ridout *et al.*, 1998; Potts & Elith, 2006). Also called the hurdle model, this approach implies first modelling the probability of presence, assuming a binomial distribution and therefore defining suitable habitats where this probability is non-null, then modelling the abundance in suitable habitats only, using either a Poisson or a truncated Poisson distribution. The former was assumed to permit the stochastic process resulting in unoccupied sites, even when habitat is suitable (Potts & Elith, 2006). This modelling approach has resulted in multiple ecological applications (e.g. Welsh *et al.*, 1996; Tyre *et al.*, 2001; Barry & Welsh, 2002).

Suppose the counts y_i with $i = 1... n$ are independent and

$y_i = 0$ with probability $1 - P(x)$

$y_i \sim$ Poisson($\lambda(z)$) with probability $P(x)$,

so that

$$P(y = 0 \mid x) = 1 - p(x) \tag{1}$$

$$P(y = r \mid x, \lambda(z)) = \frac{\lambda(z)^r \exp(-\lambda(z)) p(x)}{r!} \quad r = 0, 1, 2,... \tag{2}$$

where $\lambda(z)$ is the mean of the Poisson distribution, and then predicting the mean abundance $E(Y) = \mu(x, \lambda(z))$ can be approximated by $\mu(x, \lambda(z)) = P(x) \lambda(z)$ (Welsh *et al.*, 1996). The response (i.e. counts) is therefore modelled as having two states: a state in which no animals occur because of unsuitable habitat (equation 1) and a state in which habitat is suitable and where the response takes varying levels of abundance (*r*; equation 2), including the possibility of null abundance to accommodate the stochastic processes as mentioned earlier (Potts & Elith, 2006). In essence, the two-step hurdle model

Table 1 Candidate model predictors for abalone presence and abundance statistical models: unit, mean, standard deviation (SD) and range observed at the surveyed sites and over the prediction grid. Predictors include the mean sea surface temperature in August (*AugustSST*) and its interannual standard deviation (SD*AugSST*), the mean depth (*Depth*), the Euclidean distance to boat ramps (Dist), the mean annual harvest rate (*Harvest*) and the presence of a reef (*Reef*).

Predictor	Unit	Mean	SD	Range (sites)	Range (grid)
AugustSST	°C	13.50	0.69	10.46–15.39	10.35–17.38
SDAugSST	°C	0.41	0.10	0.18–0.85	0.12–1.27
Depth	m	8.79	6.29	0.00–30.00	0.00–30.00
Distance	°	0.15	0.18	0.00–0.89	0.00–4.49
Harvest	t year^{-1}	9.46	25.53	0.00–190.06	0.00–200.03
Reef	Binary	0.38	0.49	0–1	0–1

is similar to a mixture model using a zero-inflated Poisson (ZIP) distribution (Wenger & Freeman, 2008) that models, through a single step, (1) the probability of presences and (2) abundance conditional presence. However, decomposing the two steps through the hurdle model allows for model diagnostics at each step and, importantly, warrants the use of the same approach with other modelling techniques (e.g. machine learning algorithms) for which the ZIP distribution has not been implemented so far. Moreover, the hurdle model has the potential to outperform ZIP mixture models (Potts & Elith, 2006).

Generalized linear models

We considered a first model set combining each temperature covariate (*AugustSST*, *MarchSST* and *AnnualSST*), with or without its standard deviation from the 20-year average (SD*AugSST*, SD*MarSST* and SD*AnnSST*), with or without its quadratic term (i.e. to account for possible Gaussian response curves; Potts & Elith, 2006), to predict abalone presence (as per equation 1) and abundance conditional on presence (as per equation 2) for each species. For each response variable, all combinations of each covariate were compared based on the Akaike's information criterion corrected for small sample sizes (AIC$_c$), an index of Kullback–Leibler information loss (Burnham & Anderson, 2002). The top-ranked model according to AIC$_c$ was retained for the final model sets.

Asymptotic indices of information loss were used to assign relative strengths of evidence to the different competing models, with both AIC$_c$ and the dimension-consistent Bayesian information criterion (BIC; an approximation of the Bayes factor given no informative prior information on relative model support; Burnham & Anderson, 2002) used for model comparisons. We used BIC weights to determine the contribution of the most important variables and AIC$_c$ weights to identify the most useful predictive models given the data (Link & Barker, 2006). Weighted-average model predictions were made over the set of models considered (i.e. the sum of the predicted response variables multiplied by the model's AIC$_c$ weight) to properly account for model selection uncertainty (Burnham & Anderson, 2004; Link & Barker, 2006). The per cent deviance explained (*D*) was used as a measure of a model's goodness-of-fit. For presence models, we also considered Cohen's Kappa statistics (κ) (Cohen, 1960), which varies from ≤ 0 (for no agreement between observed and predicted probability of occurrence) to 1 (when perfect agreement is verified). Following the study by Woodby *et al.* (2009), we considered that a model had a poor performance when $\kappa < 0.4$, good when $0.4 < \kappa < 0.75$ and excellent when $\kappa > 0.75$.

Model diagnostics included examination of residual distributions based on the normal scores of standardized residual deviance (i.e. Q–Q plots) and the regression of fitted against observed values. For each response variable, the mean prediction error was assessed using a 10-fold cross-validation (Davison & Hinkley, 1997). This bootstrap resampling procedure estimates a mean prediction error for 10% of observations

that were randomly omitted from the calibration dataset; this procedure was iterated 1000 times. Spatial correlograms, assessing autocorrelation in presence and abundance (raw data and GLM residuals) as a function of the distance between sites, were generated using Moran's *I* (Diggle & Ribeiro, 2007). Evidence for spatial autocorrelation was assessed after a Bonferroni correction (Legendre & Legendre, 1998). GLM were fit in R 2.9.1 (http://www.r-project-org).

Boosted regression trees

Boosted regression trees is a machine learning algorithm that uses many simple decision trees to iteratively boost the predictive performance of the final models (Elith *et al.*, 2008). Model settings include the learning rate (lr) that controls the contribution of each tree to the final model and tree complexity (tc) that determines the extent to which interactions were fitted. The number of trees (nt) that achieved minimal predictive deviance (i.e. the loss in predictive performance because of a suboptimal model) was determined using cross-validation (function gbm.step with tc = 2, lr = 0.001 and bag fraction = 0.5; Elith *et al.*, 2008).

The same predictors as for GLM were used in models predicting the presence (as per equation 1) and abundance conditional on presence (as per equation 2) for each species. The relative contribution of the predictors to the final models was determined based on the variable importance score (%). For each response variable, the mean prediction error was assessed using a 10-fold cross-validation (see GLM section). Spatial autocorrelation in model residuals was evaluated in the same way as for GLM. BRT were fit in R 2.9.1 (R Development Core Team 2009) using package GBM, tutorial and functions provided by Elith *et al.* (2008).

Out-of-sample validation

For both GLM and BRT, we used the South Australian dataset (including abundance records of both species) to fit the models and the Victorian dataset (including only abundance records of *H. rubra*, because *H. laevigata* is absent or in low abundances in Victoria) as an independent dataset for the out-of-sample validation, which was therefore performed on *H. rubra* only because of restricted data availability. The best supported GLM and BRT for *H. rubra* were used to predict abalone abundance at each of the sites surveyed in Victoria (i.e. tested on an independent dataset).

Multimodel ensemble predictions

We computed model predictions across the entire study area for both the GLM and BRT (it is noted that, because data were unavailable in Victoria for *H. laevigata*, models were fitted on the South Australian dataset only, to keep a consistent methodology for both species). Predicted abalone grid cell abundance for each technique was weighted in proportion to the model structural deviance explained by the training data

(scaled between 0 and 1) and then averaged to provide a weighted-average model predictions of abalone abundance. In each case, we computed the mean prediction error (in % of the observed value) and the regression of predicted against observed abundances. The resulting predictive maps were qualitatively assessed by an expert of abalone ecology and biology (S.A. Shepherd).

RESULTS

Generalized linear models

Generalized linear models of temperature covariates showed that *AugustSST* and SD*AugSST* provided the best temperature model for *H. rubra* (wAIC$_c$ = 1) when predicting presence and abundance conditional on presence, whereas *AugustSST*, its quadratic term and SD*AugSST* provided the best model for *H. laevigata* (wAIC$_c$ = 1) when predicting presence and abundance conditional on presence (Table 1).

For *H. laevigata*, the *Reef* model outperformed the null only when predicting presence (as all positive abundance records of this species were found on reefs), whereas for both abalone species, *Harvest* outperformed the null only when predicting abundance (Table 2). Of all combinations of predictors, only those that outperformed the null model and with an AIC$_c$ weight > 10^{-3} were retained. Therefore, for both species, the final presence model sets gathered the predictors *Reef* and a combination of temperature covariates (as determined earlier), *Depth* and *Distance*. The final abundance model sets consisted of a combination of *Harvest*, temperature covariates (as determined earlier), *Depth* and *Distance*. *Reef* was also included as a predictor of *H. rubra* abundance.

For both *H. rubra* and *H. laevigata*, *Reef* combined with *AugustSST* (plus its quadratic term for *H. laevigata*) provided

the best-fitting presence GLM according to AIC$_c$, explaining 33.7 and 27.2% of structural model deviance, respectively (Tables 2 and 3). In both cases, there was a good agreement between fitted and observed values from the South Australian dataset (κ = 0.48 and 0.46). The saturated model provided the best-fitting model of abundance conditional on presence according to AIC$_c$ for both species (accounting for 20.7% and 38.9% of deviance in abundance, respectively; Tables 2 and 3), followed by the *Harvest* model (plus *Reef* for *H. rubra* only). In most cases, we found agreement between AIC$_c$ and BIC rankings.

Generalized linear models residuals were normally distributed, and when the predicted probability of presence and the abundance conditional on presence were combined, the regression of fitted against observed abundance values returned a coefficient of determination R^2 of 0.53 for *H. rubra* and 0.56 for *H. laevigata* (see Fig. S2). However, a regression slope < 1 indicated that fitted values were generally lower than observed ones, showing the limited ability of the GLM to predict high abundance values. The 10-fold cross-validation indicated that the mean prediction error when predicting the probability of presence was 0.28 for *H. rubra* and 0.36 for *H. laevigata*. The mean prediction error when predicting the abundance conditional on presence was 37.3% for *H. rubra* and 46.7% for *H. laevigata*. No evidence of spatial autocorrelation was found in GLM residuals (i.e. $P > 0.05$ at all distance classes and for all models).

Boosted regression trees

The optimal number of trees for presence BRT was 6750 and 2500 for *H. rubra* and *H. laevigata*, respectively, with the resulting models accounting for 37% and 29% of the deviance in the probability of their presence. *Reef* and *AugustSST* were

Table 2 *Haliotis rubra* model fits: summary of the generalized linear models predicting the probability of presence (P) and the abundance conditional on presence (N) for *H. rubra*. Model predictors include the presence of a reef (binary; *Reef*), the mean sea surface temperature in August (°C; *AugustSST*) and its interannual standard deviation (°C; SD*AugustSST*), the mean depth (m; *Depth*), the Euclidean distance to boat ramps (°; *Dist*) and the mean annual harvest rate (t year^{-1}).

Model	k	LL	wAIC$_c$	wBIC	D	κ
$P \sim$ Reef + AugustSST	3	−88.02	0.46	0.78	33.7	0.48
$P \sim$ Reef + Depth + AugustSST	4	−87.63	0.23	0.08	34.0	0.40
$P \sim$ Reef + AugustSST + SDAugSST	4	−88.02	0.16	0.06	33.7	0.46
$P \sim$ Reef + Depth + Dist + AugustSST + SDAugSST	6	−86.43	0.09	0.00	34.9	0.35
$P \sim$ Reef + Depth	3	−90.40	0.04	0.07	31.9	0.36
$P \sim$ Reef + Depth + Dist	4	−90.38	0.02	0.01	31.9	0.36
$P \sim 1$	1	−132.75	0.00	0.00	0.0	0.00
$N \sim$ reef + depth + Dist + AugustSST + SDAugSST + harvest	7	−2127.04	1.00	1.00	20.7	
$N \sim$ Reef + Harvest	3	−2244.66	0.00	0.00	16.3	
$N \sim$ Reef + Depth + Dist	4	−2419.32	0.00	0.00	9.8	
$N \sim$ Reef + AugustSST + SDAugSST	4	−2441.69	0.00	0.00	8.9	
$N \sim 1$	1	−2681.71	0.00	0.00	0.0	

With k, number of parameters; LL, log likelihood; wAIC$_c$, weight given by the Akaike's information criterion corrected for small sample size; wBIC, weight given by the Bayesian information criterion; D, per cent deviance explained; κ, Cohen's Kappa statistics.

Table 3 *Haliotis laevigata* model fits: summary of the generalized linear models predicting the probability of presence (*P*) and the abundance conditional on presence (*N*) for *H. laevigata*. Model predictors include the presence of a reef (binary; *Reef*), the mean sea surface temperature in August (°C; *AugustSST*), its interannual standard deviation (°C; *SDAugSST*) and quadratic term (*AugSST^2*), the mean depth (m; *Depth*), the Euclidean distance to boat ramps (°; *Dist*) and the mean annual harvest rate (t year^{-1}).

Model	k	LL	wAIC$_c$	wBIC	D	κ
$P \sim$ Reef + AugustSST + AugSST^2	4	−101.78	0.42	0.68	27.2	0.46
$P \sim$ Reef + AugustSST + AugSST^2 + SDAugSST	5	−101.29	0.24	0.08	27.6	0.40
$P \sim$ Reef + Depth + AugustSST + AugSST^2	5	−101.39	0.22	0.07	27.5	0.42
$P \sim$ Reef + Depth + Dist + AugustSST + AugSST^2 + SDAugSST	7	−100.25	0.08	0.00	28.3	0.41
$P \sim$ Reef + Depth + Dist	4	−104.37	0.03	0.05	25.3	0.36
$P \sim$ Reef + Depth	3	−106.22	0.01	0.11	24.0	0.36
$P \sim 1$	1	−139.82	0.00	0.00	0.0	0.00
$N \sim$ Depth + Dist + AugustSST + AugSST^2 + SDAugSST + Harvest	6	−1100.97	1.00	1.00	38.9	
$N \sim$ Harvest	2	−1308.16	0.00	0.00	27.4	
$N \sim$ AugustSST + AugSST^2 + SDAugSST	3	−1438.35	0.00	0.00	20.2	
$N \sim$ Depth + Dist	3	−1477.77	0.00	0.00	18.0	
$N \sim 1$	1	−1802.03	0.00	0.00	0.0	

With *k*, number of parameters; LL, log likelihood; *w*AIC$_c$, weight given by the Akaike's information criterion corrected for small sample size; *w*BIC, weight given by the Bayesian information criterion; *D*, per cent deviance explained; κ, Cohen's Kappa statistics.

the most important predictors of abalone presence for both species, although in a different order (Fig. 2). The 10-fold cross-validation returned a mean prediction error with respect to the probability of presence of 0.29 for *H. rubra* and 0.37 for *H. laevigata*. The optimal number of trees for BRT predicting abundance conditional on presence was 2250 and 1450 for *H. rubra* and *H. laevigata*, respectively, with the resulting models accounting for 55% and 45% of the deviance in abundance conditional on presence. The most important predictors of abundance conditional on presence were *August-SST* and SD*AugSST* for *H. rubra*, and *Depth* and *Distance* for *H. laevigata* (Fig. 2). The 10-fold cross-validation returned a mean prediction error (in % of the observed abundance) of 45% for *H. rubra* and 37% for *H. laevigata*.

Out-of-sample validation

The out-of-sample validation of *H. rubra* models using the Victoria abundance dataset returned a mean prediction error of 26.4% for GLM, 39.2% for BRT and 30.9% for the multimodel ensemble approach. The regression of predictions against observations showed that the highest (adjusted) R^2 was obtained for the multimodel ensemble approach (0.16), followed by GLM (0.12) and BRT (0.06; see Fig. S3).

Multimodel ensemble predictions

Predictive maps showed similar abundance patterns for both species (Fig. 3), decreasing from the western to the eastern limits of the study area. Maximal abundance values were reached on outer reefs, especially for *H. laevigata* for which only low abundances were predicted inside the two South Australian gulfs. Predicted abundances over the grid were generally lower in Victoria than in South Australia (mean

abundance per 100 m^2 = 6 and 12 individuals, respectively) although, at least for *H. rubra*, abalone were locally more abundant on reefs surveyed in Victoria than in South Australia (mean abundance per 100 m^2 = 43 and 11 individuals, respectively).

DISCUSSION

Our approach showed that it is possible to model the occurrence and abundance patterns of marine invertebrates using key environmental drivers (albeit with a relatively high degree of unexplained variation) and use these models to build predictive maps of species abundance at scales useful to management (*c.* 1 km^2). This approach, developed across a broad geographic range, and validated using an independent dataset for one of the two species, provides an example of where SDMs can deliver useful support for the management of commercially important marine species, by quantifying spatial patterns of habitat suitability. To date, the stock management of *H. rubra* and *H. laevigata* has been built on outputs from non-spatial stochastic and deterministic population modelling techniques (e.g. Bardos *et al.*, 2006). Our spatial modelling provides a key understanding of spatial patterns of abalone abundance and their primary drivers. Coupling our spatially explicit information about species range and habitat with demographically explicit population modelling approaches will strengthen conservation planning (Brook *et al.*, 2009), ensuring the long-term persistence of abalone stocks in southern Australia.

Both machine learning and/or statistical modelling methods used in this study (i.e. GLM and BRT) selected similar climate and environmental drivers of distribution for both species, including the presence of reefs and August SST as major drivers of abalone occurrence and abundance, respectively.
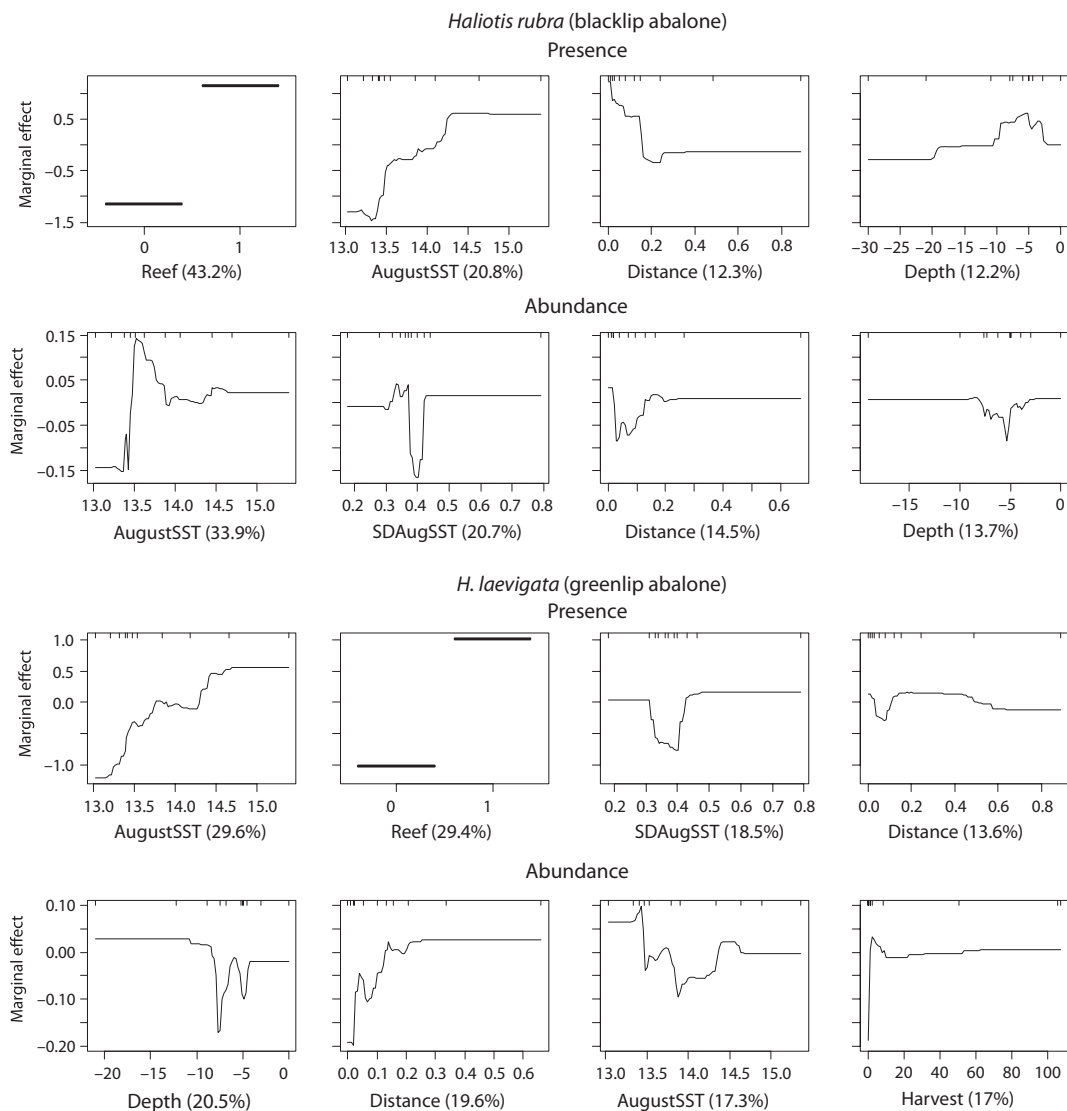
Figure 2 Partial dependence plots for the four most influent variables in the boosted regression trees predicting the probability of presence and abundance of *Haliotis rubra* and *H. laevigata*. See Table 1 for explanation of variables, units and descriptive statistics. *Y*-axes are on the logit scale and are centred to have zero mean over the data distribution. Rug (bivariate scatter) plots at inside top of plots show distribution of sites across that variable, in deciles.

These results agree with previous, empirical studies identifying the major determinants of abalone distribution (Shepherd, 1973). However, GLM and BRT differed in their ability to identify species-specific predictors of abundance. For example, BRT identified depth as a predictor of *H. laevigata*, but GLM did not. This difference might reflect the ability of BRT to accommodate for complex, nonlinear species–environment relationships that are likely to differ subtly between species. Conversely, GLM can usually only fit linear, quadric or cubic relationships (Elith *et al.*, 2006); however, one of their advantages is that they are less prone to overfitting than BRT (Elith *et al.*, 2008).

Model goodness-of-fit was similar for GLM and BRT, with respectively 21–39% and 29–45% of deviance accounted for in the presence or abundance of either species. These levels of

goodness-of-fit are consistent with the few existing distribution models of marine molluscs, with GLM explaining 42.7% of deviance in scallop abundance patterns in Uruguay (59% when including quadratic terms; Gutierrez *et al.*, 2008) and 6.2–26.7% of deviance in bivalve mesoscale occurrence patterns in New Zealand (16–52% when including microhabitat characteristics; Inglis *et al.*, 2006). In freshwater ecosystems of the UK, spatial logistic regression explained up to 34% of deviance in mollusc occurrence (Sanderson *et al.*, 2005). For all these models, a single modelling technique was used, without validating the subsequent model on an independent dataset. By combining GLM and BRT, our ensemble predictions outperformed individual models when applied to independent data and showed reasonably good model transferability, despite low proportions of variance in the validation dataset
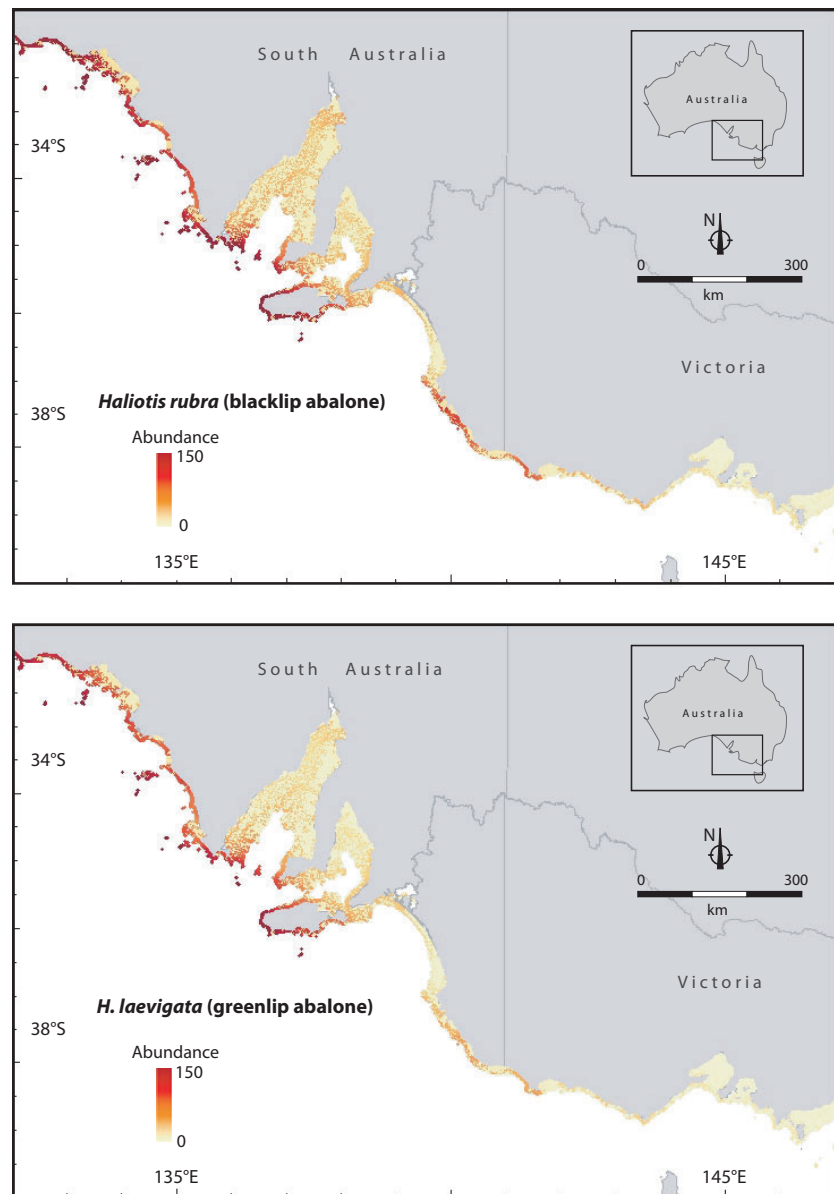
Figure 3 Multimodel ensemble predicted abundances (estimated within a 100-m² survey area) of *Haliotis rubra* and *H. laevigata* in coastal habitats of South Australia and Victoria between 0 and 30 m depth (projection: Geocentric Datum of Australia 1994).

explained by the models (i.e. R^2). The validation dataset showed a large inherent variability that is likely to result from variation in microhabitat characteristics, which could not be accounted for by the models because of data paucity.

An expert assessment of the final predictive maps found that patterns of relative abalone abundance showed a high degree of accuracy across the study area, especially for *H. rubra*, albeit slight overestimations in some locations (S.A. Shepherd, pers. comm.). For *H. rubra*, a substantial component of this overestimation reflects the effects of overfishing, where the predicted values appear to reflect reasonable estimates of abundance before exploitation. For *H. laevigata*, however, positive densities were predicted in areas where the species does not occur. For example, this species has never been observed on the inshore reefs in the far western part of the study area, while it occurs on the offshore islands (S.A.

Shepherd, pers. comm.). Our predictions of high abundances on the inshore reefs are probably due to the *Reef* predictor that does not distinguish between the different microhabitat requirements of each species. Indeed, *H. laevigata* tends to be found on flatter, open reef, while *H. rubra* tends to be found in crevices on topographically complex reef (Shepherd, 1973). Therefore, developing improved spatial layers of reef microhabitat would possibly strengthen predictions of abundance patterns across the species' range. In contrast, in several locations where high abundances were predicted, *H. laevigata* have been historically observed but no longer occupy these areas (e.g. north-eastern side of Kangaroo Island; longitude 136.5–136.7° E; latitude 35.7–35.9° S). Historical events, in particular overfishing, have caused population depletion in suitable habitats (S.A. Shepherd, pers. comm.) and therefore may have been responsible for differences between distribution

patterns based on the fundamental niche (which our models more closely predict) and realized distribution (Hirzel & Le Lay, 2008). Because of limited data availability, our predictors *Harvest* and *Distance* did not incorporate historical or fine-scale fishing pressure and therefore failed to account for potential overfishing effects in the past and areas heavily harvested within fishing zones. This highlights another advantage of SDMs, that is, their potential usefulness to identify suitable habitats for species recolonization (Cianfrani *et al.*, 2010).

We suspect that model performance could have been even higher with the addition of other covariates that are relevant to the species' ecology. For instance, water movement influences the availability of drifting algae, affecting the feeding of abalone and their fine-scale distribution patterns within reefs (Shepherd, 1973). However, fine-scale 3D ocean hydrodynamics data are not available in most locations, with the best model simulations for the study area being available at a 0.1°

(*c.* 10 km) horizontal resolution and with a depth resolution of 10 m (Oke *et al.*, 2008), which at present make them unsuitable for the purpose of our study. Even in static SDMs, variables that describe average hydrodynamic conditions such as average current speed or bottom stress could provide useful predictors. Given that worldwide ocean hydrodynamics are predicted to change in strength and/or direction under climate change, altering food availability and dispersal patterns for many marine species (e.g. Ruhl & Smith, 2004; Munday *et al.*, 2009), there is an urgent need to generate fine-resolution ocean hydrodynamic models that integrate forecasts of future climate change and integrate this spatial information in SDMs. This absence of hydrodynamic data reflects a broader issue in distribution modelling: missing covariates, which can result in spurious predictions or spatial autocorrelation in residuals (Barry & Elith, 2006). One must bear in mind the uncertainty that propagates through the different modelling steps, as well as the variation that is not explained by the models and could
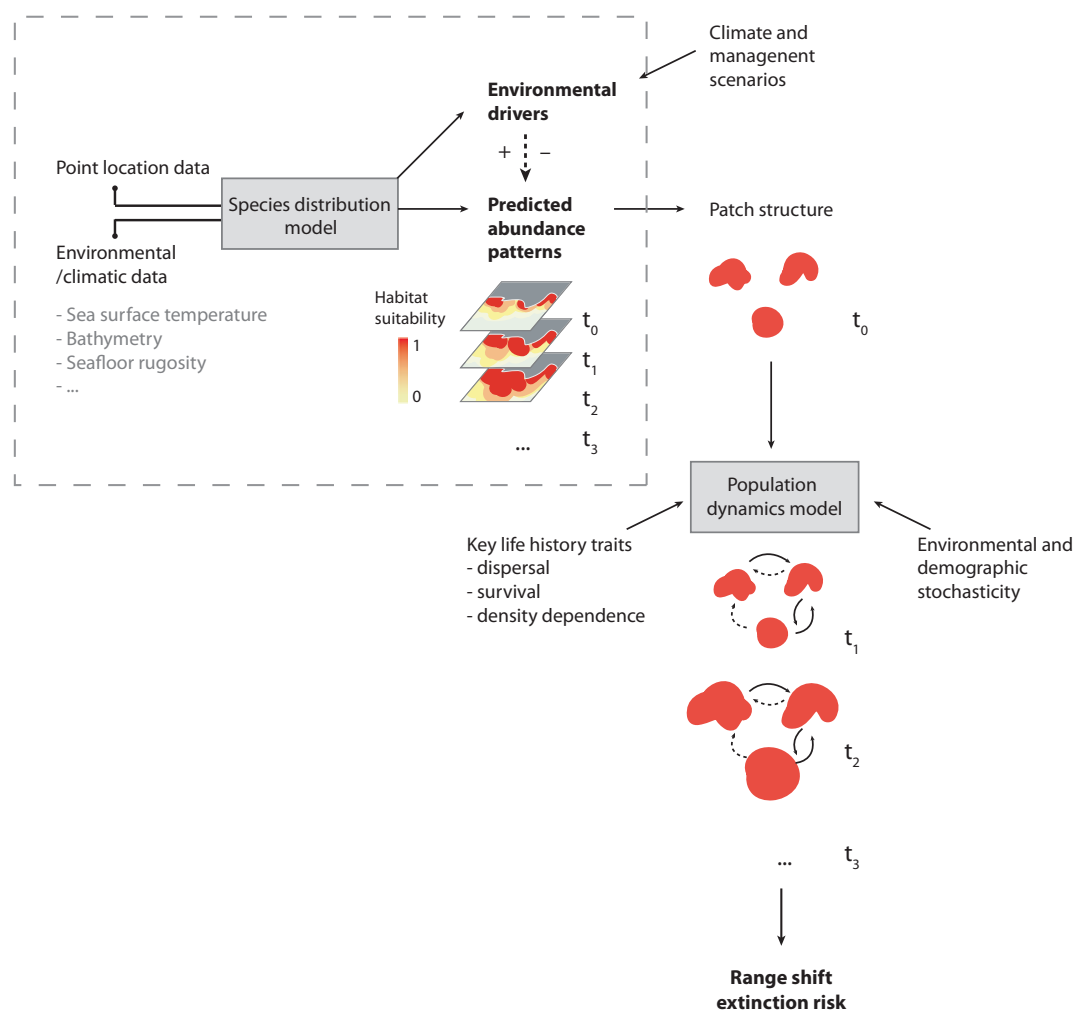

Figure 4 Schematic diagram of a coupled spatio-demographic modelling approach in marine systems. The dotted rectangle outlines the modelling steps developed in this study. As a next step, key spatial drivers of habitat suitability will be used to define spatial patch structure and how this is expected to change in response to annual predictions of future environmental conditions. The population dynamics model is used to capture some of the complexities and uncertainties underlying biological mechanisms that drive *H. rubra* and *H. laevigata* distribution and abundance in response to forecasts of future climate change and harvest pressure (see Discussion for further details).

be linked to factors such as measurement error or missing covariates (e.g. Barry & Elith, 2006). In many ways, overlooked or unavailable predictor information can be viewed as a prime factor limiting the predictive power of SDMs, including those developed in this study, especially when forecasting into novel environments. More recently, species traits have been shown to influence the spatial and temporal transferability of SDM projections (Syphard & Franklin, 2010; Dobrowski *et al.*, 2011). Because most SDMs are correlative, not mechanistic, they do not link demographic traits (e.g. population growth, dispersal and trending vital rates) and ecosystem processes (i.e. ecological interactions and feedbacks) explicitly to an organism's perception and use of spatial features. Consequently, when SDMs are transferred through time or across geographic regions, they provide only a first approximation of spatial patterns of habitat suitability (Pearson & Dawson, 2003).

Defining the key drivers of present-day spatial abundance is the first step towards better understanding future impacts of global change on species persistence (Fordham *et al.*, in press). As the ecological traits that make *H. rubra* and *H. laevigata* particularly amenable to SDM approaches also make them susceptible to rapid environmental change (see Introduction), and SST (a key predictor of abundance) is expected to continue to warm because of anthropogenic climate change (Clement *et al.*, 2010; Xie *et al.*, 2010), the long-term management of these species relies on informed future projections. By linking SDM forecasts to structured spatial population models, which are able to integrate seascape features and demographic and ecosystem processes, a more biologically realistic assessment of extinction risk under climate change can be made (Keith *et al.*, 2008; Brook *et al.*, 2009). Accordingly, we are now developing a model framework to capture some of the complexities and uncertainties underlying biological mechanisms driving *H. rubra* and *H. laevigata* distribution and abundance (Fig. 4), which should provide a better diagnosis of the potential threat that global climate change, either alone or in combination with other human-mediated impacts, poses to abalone stocks in southern Australia.

The Australian abalone industry is worth more than $100 million annually. The governance of the South Australian component ($30 million per annum) is split among three independent sectors of production, management and government lobby. It is unlikely that the economic future of these sectors will be evenly affected by future climate and fishing. Such forecasts of future abundance therefore will be useful to current and future planning within this industry, particularly to those sectors that need to balance conservation actions (e.g. harvest quotas and marine reserves), and sustainable aquaculture, against concomitant stock decline caused by changing climate.

ACKNOWLEDGEMENTS

REFERENCES

Araujo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Araujo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.

Bardos, D.C., Day, R.W., Lawson, N.T. & Linacre, N.A. (2006) Dynamical response to fishing varies with compensatory mechanism: an abalone population model. *Ecological Modelling*, **192**, 523–542.

Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.

Barry, S.C. & Welsh, A.H. (2002) Generalized additive modelling and zero inflated count data. *Ecological Modelling*, **157**, 179–188.

Berg, M.P., Kiers, E.T., Driessen, G., van der Heijden, M., Kooi, B.W., Kuenen, F., Liefting, M., Verhoef, H.A. & Ellers, J. (2010) Adapt or disperse: understanding species persistence in a changing world. *Global Change Biology*, **16**, 587–598.

Botkin, D.B., Saxe, H., Araujo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T., Chesson, P., Dawson, T.P., Etterson, J.R., Faith, D.P., Ferrier, S., Guisan, A., Hansen, A.S., Hilbert, D.W., Loehle, C., Margules, C., New, M., Sobel, M.J. & Stockwell, D.R.B. (2007) Forecasting the effects of global warming on biodiversity. *BioScience*, **57**, 227–236.

Brook, B.W., Akçakaya, H.R., Keith, D.A., Mace, G.M., Pearson, R.G. & Araujo, M.B. (2009) Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. *Biology Letters*, **5**, 723–725.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information theoretic approach*, 2nd edn. Springer-Verlag, New York.

Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference – understanding AIC and BIC in model selection. *Sociological Methods and Research*, **33**, 261–304.

Chick, R.C., Turich, N., Mayfield, S. & Dent, J. (2009) Western zone abalone (*Haliotis rubra & H. laevigata*) fishery (Region A), *Fishery assessment report for PIRSA* SARDI Aquatic Sciences, Adelaide.

Cianfrani, C., Le Lay, G., Hirzel, A.H. & Loy, A. (2010) Do habitat suitability models reliably predict the recovery areas of threatened species? *Journal of Applied Ecology*, **47**, 421–430.

Clement, A.C., Baker, A.C. & Leloup, J. (2010) Climate change: patterns of tropical warming. *Nature Geoscience*, **3**, 8–9.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Eductional and Psychological Measurement*, **20**, 37–46.

Davison, A.C. & Hinkley, D.V. (1997) *Bootstrap methods and their application*. Cambridge University Press, Cambridge.

Diggle, P.J. & Ribeiro, P.J. Jr. (2007) *Model-based geostatistics*. Springer, New York.

Dobrowski, S., Thorne, J., Greenberg, J., Safford, H., Mynsberge, A., Crimmins, S. & Swanson, A. (2011) Modeling plant ranges over 75 years of climate change in California, USA: relating transferability to species traits. *Ecological Monographs*, **81**, 241–257.

EconSearch (2010) *Economic Indicators for the South Australian abalone fishery 2008/09*. Report prepared for Primary Industries and Resources, Adelaide, South Australia.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

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

Fordham, D.A., Akçakaya, H.R., Araujo, M.B. & Brook, B.W. (in press) Modelling range shifts for invasive vertebrates in response to climate change. *Wildlife conservation in a changing climate* (ed. by J. Brodie, E. Post and D. Doak), University of Chicago Press, Chicago.

Gaston, K.J. & Fuller, R.A. (2009) The sizes of species' geographic ranges. *Journal of Applied Ecology*, **46**, 1–9.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.

Gutierrez, N.L., Martinez, A. & Defeo, O. (2008) Identifying environmental constraints at the edge of a species' range: scallop Psychrochlamys patagonica in the SW Atlantic Ocean. *Marine Ecology-Progress Series*, **353**, 147–156.

Hirzel, A.H. & Le Lay, G. (2008) Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45**, 1372–1381.

Hutchinson, M.F. (1995) Interpolating mean rainfall using thin-plate smoothing splines. *International Journal of Geographical Information Systems*, **9**, 385–403.

Hutchinson, M. (2006) *Anusplin version 4.36*. Centre for Resource and Environmental Studies. The Australian National University, Canberra, Australia.

Inglis, G.J., Hurren, H., Oldman, J. & Haskew, R. (2006) Using habitat suitability index and particle dispersion models for early detection of marine invaders. *Ecological Applications*, **16**, 1377–1390.

Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Keith, D.A., Akcakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araujo, M.B. & Rebelo, T.G. (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, **4**, 560–563.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*, 2nd English edn. Elsevier, Amsterdam. pp. 1–853.

Ling, S.D., Johnson, C.R., Ridgway, K., Hobday, A.J. & Haddon, M. (2009) Climate-driven range extension of a sea urchin: inferring future trends by analysis of recent population dynamics. *Global Change Biology*, **15**, 719–731.

Link, W.A. & Barker, R.J. (2006) Model weights and the foundations of multimodel inference. *Ecology*, **87**, 2626–2635.

Marullo, S., Nardelli, B.B., Guarracino, M. & Santoleri, R. (2007) Observing the Mediterranean Sea from space: 21 years of Pathfinder-AVHRR sea surface temperatures (1985 to 2005): re-analysis and validation. *Ocean Science*, **3**, 299–310.

Mayfield, S., Carlson, I.J. & Chick, R.C. (2008) Central zone abalone fishery (*Haliotis laevigata* and *H. rubra*), *Fishery assessment report to PIRSA Fisheries* p. 71. SARDI Aquatic Sciences, Adelaide.

Mayfield, S., Hogg, A., Saunders, T.M. & Burch, P. (2009) Southern zone abalone fishery (*Haliotis rubra and H. laevigata*), *Fishery assessment report for PIRSA*. SARDI Aquatic Sciences, Adelaide.

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

Mellin, C., Andréfouët, S., Kulbicki, M., Dalleau, M. & Vigliola, L. (2009) Remote sensing and fish-habitat relationships in coral reef ecosystems: review and pathways for systematic multi-scale hierarchical research. *Marine Pollution Bulletin*, **58**, 11–19.

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.

Munday, P.L., Leis, J.M., Lough, J.M., Paris, C.B., Kingsford, M.J., Berumen, M.L. & Lambrechts, J. (2009) Climate change and coral reef connectivity. *Coral Reefs*, **28**, 379–395.

Nobles, M., Casement, D. & Mayfield, S. (2004) Management plan for the South Australian abalone fishery, *The South Australian fisheries management series, paper no. 42* p. 47. Primary Industries and Resources of South Australia, Adelaide.

Oke, P.R., Brassington, G.B., Griffin, D.A. & Schiller, A. (2008) The Bluelink ocean data assimilation system (BODAS). *Ocean Modelling*, **21**, 46–70.

Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

Poloczanska, E.S., Hawkins, S.J., Southward, A.J. & Burrows, M.T. (2008) Modeling the response of populations of competing species to climate change. *Ecology*, **89**, 3138–3149.

Potts, J.M. & Elith, J. (2006) Comparing species abundance models. *Ecological Modelling*, **199**, 153–163.

Prince, J.D., Sellers, T.L., Ford, W.B. & Talbot, S.R. (1987) Experimental-evidence for limited dispersal of haliotid larvae (Genus Haliotis, Mollusca, Gastropoda). *Journal of Experimental Marine Biology and Ecology*, **106**, 243–263.

R Development Core Team (2009) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Ridout, M., Demetrio, C. & Hinde, J. (1998) Models for count data with may zeros. *International Biometric Conference, Cape Town* 1–130.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P. & Richardson, A.J. (in press) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2010.00636.x.

Ruhl, H.A. & Smith, K.L. (2004) Shifts in deep-sea community structure linked to climate and food supply. *Science*, **305**, 513–515.

Sanderson, R.A., Eyre, M.D. & Rushton, S.P. (2005) Distribution of selected macroinvertebrates in a mosaic of temporary and permanent freshwater ponds as explained by autologistic models. *Ecography*, **28**, 355–362.

Saunders, T., Mayfield, S. & Hogg, A. (2009a) Using a simple morphometric marker to identify spatial units for abalone fishery management. *ICES Journal of Marine Science*, **66**, 305–314.

Saunders, T.M., Connell, S.D. & Mayfield, S. (2009b) Differences in abalone growth and morphology between locations with high and low food availability: morphologically fixed or plastic traits? *Marine Biology*, **156**, 1255–1263.

Shepherd, S.A. (1973) Studies on Southern Australian Abalone (Genus-Haliotis). 1. Ecology of 5 Sympatric Species. *Australian Journal of Marine and Freshwater Research*, **24**, 217–257.

Statzner, B. & Beche, L.A. (2010) Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology*, **55**, 80–119.

Syphard, A.D. & Franklin, J. (2010) Species traits affect the performance of species distribution models for plants in southern California. *Journal of Vegetation Science*, **21**, 177–189.

Temby, N., Miller, K. & Mundy, C. (2007) Evidence of genetic subdivision among populations of blacklip abalone (*Haliotis rubra* Leach) in Tasmania. *Marine and Freshwater Research*, **58**, 733–742.

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F. & De Clerck, O. (2011) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2011.00656.x.

Tyre, A.J., Possingham, H.P. & Lindenmayer, D.B. (2001) Inferring process from pattern: can territory occupancy provide information about life history parameters? *Ecological Applications*, **11**, 1722–1737.

Vandewalle, M., de Bello, F., Berg, M.P., Bolger, T., Doledec, S., Dubs, F., Feld, C.K., Harrington, R., Harrison, P.A., Lavorel, S., da Silva, P.M., Moretti, M., Niemela, J., Santos, P., Sattler, T., Sousa, J.P., Sykes, M.T., Vanbergen, A.J. & Woodcock, B.A. (2010) Functional traits as indicators of biodiversity response to land use changes across ecosystems and organisms. *Biodiversity and Conservation*, **19**, 2921–2947.

Wahba, G. (1990) Spline models for observation data, *CBMS-NSF Regional Conference Series in Applied Mathematics 59* SIAM, Philadelphia, Pennsylvania.

Walton, C.C., Pichel, W.G., Sapper, J.F. & May, D.A. (1998) The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research-Oceans*, **103**, 27999–28012.

Watts, M.J., Li, Y., Russel, B.R., Mellin, C., Connell, S.D. & Fordham, D.A. (2011) A novel method for mapping reefs and subtidal rocky habitats using artificial neural networks. *Ecological Modelling*, doi:10.1016/j.ecolmodel.2011.04.024.

Webster, M.A. & Petkovic, P. (2005) Australian bathymetry and topography grid, June 2005. Geoscience Australia Record, 2005/12, 12pp.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F. & Lindenmayer, D.B. (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, **88**, 297–308.

Wenger, S.J. & Freeman, M.C. (2008) Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology*, **89**, 2953–2959.

Woodby, D., Carlile, D. & Hulbert, L. (2009) Predictive modeling of coral distribution in the Central Aleutian Islands, USA. *Marine Ecology-Progress Series*, **397**, 227–240.

Xie, S.P., Deser, C., Vecchi, G.A., Ma, J., Teng, H.Y. & Wittenberg, A.T. (2010) Global warming pattern formation: sea surface temperature and rainfall. *Journal of Climate*, **23**, 966–986.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Spatial layers available across the study area for model predictors.

Figure S2 Diagnostic plots for the generalized linear models predicting the abundance of *Haliotis rubra* and *H. laevigata*.

Figure S3 Out-of-sample validation: regression of predicted against observed values of *Haliotis rubra* abundance in Victoria.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

Editor: Reuben Keller