



Selection of predictor variables for species distribution models: a case study with an invasive marine bryozoan

Conrad James Pratt¹ · Danielle Denley² · Anna Metaxas¹

Received: 7 May 2021 / Accepted: 9 January 2022 / Published online: 26 January 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Species distribution models (SDMs) are important tools for predicting the occurrence and abundance of organisms in space and time, with numerous applications in ecology. However, the accuracy and utility of SDMs can be compromised when predictor variables are selected without careful consideration of their ecophysiological relevance to the focal organism. We conducted an in-depth examination of the variable selection process by evaluating predictors to be used in SDMs for *Membranipora membranacea*, an ecologically significant marine invasive species with a complex lifecycle, as a case study. Using an information-theoretic and multi-model inference approach based on generalized linear mixed models, we assessed multiple environmental variables (depth, kelp density, kelp substrate, temperature, and wave exposure) as predictors of the abundance of multiple life stages of *M. membranacea*, investigating species-environment relationships and relative and absolute variable importance. We found that the relative importance of a predictor, the metric calculated to represent a predictor, and whether a predictor was proximal or distal were important considerations in the variable selection process. Data constraints (e.g. sample size, characteristics of available predictor data) may inhibit accurate assessment of predictor variables during variable selection. Importantly, our results suggest that species-environment relationships derived from small-scale studies can inform variable selection for SDMs at larger spatiotemporal scales. We developed a conceptual framework for variable selection for SDMs which can be applied to most contexts of species distribution modelling, but particularly those with several candidate predictors and a large dataset.

Keywords Invasive species · *Membranipora membranacea* · Environmental predictors · Species distribution modelling · Kelp beds

Introduction

Correlative species distribution models (SDMs) generate relationships between data on species occurrence or abundance and a set of predictor variables to predict distribution in space and time (Elith and Leathwick 2009; Guisan et al. 2017). With broad accessibility of data (Guisan et al. 2017) and software (Jarnevich et al. 2015), species distribution modelling has become a widely used tool for a

variety of applications, including testing evolutionary and biogeographical theories, predicting shifts in the distribution of organisms under climate change, and supporting the management of invasive species (Guisan and Thuiller 2005; Franklin 2013; Guisan et al. 2017). However, despite their ease of application, the utility of SDMs relies on model accuracy (Pearson 2007). The accuracy of SDMs can be affected by many factors, including data quality (Robinson et al. 2011), the appropriateness of the applied model (Jarnevich et al. 2015), and the context in which the model is applied (e.g. extent of extrapolation; Elith and Leathwick 2009; Heikkinen et al. 2012). An important factor, often overlooked, is the selection and use of appropriate predictor variables (Elith and Leathwick 2009; Austin and Van Niel 2011; Synes and Osborne 2011; Barbet-Massin and Jetz 2014; Fourcade et al. 2018).

The selection of predictor variables with ecophysiological relevance to the modelled organism ensures the accurate

Communicated by Donald DeAngelis.

✉ Conrad James Pratt
c.pratt@dal.ca

¹ Department of Oceanography, Dalhousie University, Halifax, NS B3H 4R2, Canada

² School of Resource and Environmental Management, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

characterisation of the realised niche of the species (Peterson et al. 2011), optimising SDM accuracy and transferability (Petitpierre et al. 2017). In contrast, the accuracy and utility of SDMs can be compromised when predictors are selected without ecological justification (Austin 2007; Elith and Leathwick 2009). Examples include the use of variables because they are readily available rather than biologically linked to the modelled organism (Fourcade et al. 2018), assumption and fitting of linear relationships for predictors which may affect species distribution non-linearly (Austin 2007; Santika and Hutchinson 2009), and the use of predictors at inappropriate spatial (Dormann 2007; Peterson et al. 2011) or temporal (Peterson et al. 2011; Mod et al. 2016; Gardner et al. 2019) scales or in terms of an inappropriate metric for continuous predictors (DeWeber and Wagner 2018). Inadequate knowledge of variable importance may also impair the selection of relevant and informative predictor variables (Austin and Van Niel 2011). Additionally, for organisms with complex life cycles (*sensu* Wilbur 1980), the predictive performance and management applicability of SDMs can be diminished if they do not consider multiple life stages, which may each respond to different predictors (Costa et al. 2015; Asch and Erisman 2018). Therefore, variable selection involving a detailed investigation of the relationships between the focal species and its potential predictors, based on existing knowledge and ecological theory, should be undertaken prior to the construction of an SDM (Elith and Leathwick 2009; Fourcade et al. 2018).

Although past studies have provided guidance on various aspects of variable selection (effects of variable selection on SDM projections; Porfirio et al. 2014; selecting variables to maximize transferability; Petitpierre et al. 2017; the importance of ad hoc variable selection; Fourcade et al. 2018; representativeness of proximal predictors by distal bioclimatic variables; Gardner et al. 2019; the effect of using field-based vs. remotely-sensed predictors; La Marca et al. 2019; trade-offs between using ecologically relevant predictors and spatiotemporal proxy variables; Brodie et al. 2020), we are not aware of a study to date providing a unifying framework for variable selection for species distribution models. Williams et al. (2012) provided a framework for variable selection for “biodiversity models”, which included SDMs in its scope, but their study lacks a discussion of some of the variable selection issues outlined above (e.g. metrics for continuous predictors) due to its generality. Here, we provide an example of a rigorous, biologically informed process of variable selection using a well-studied and ecologically significant marine invasive species (*Membranipora membranacea*) with a complex life cycle. We use our results to construct a generalizable framework for variable selection for SDMs.

M. membranacea is a widely distributed, invasive bryozoan inhabiting kelp bed ecosystems in the northwest Atlantic Ocean. Native to Europe and the western coast of

North America (Schwaninger 2008), the bryozoan was first observed in the northwest Atlantic in the Gulf of Maine in 1987 (Berman et al. 1992) and has since spread to occupy a large range including Nova Scotia (Watanabe et al. 2010), coastal Newfoundland and southern Labrador (DFO 2011; Caines and Gagnon 2012), and the Gulf of St. Lawrence (Denley et al. 2019b). The bryozoan has an annual, complex lifecycle in which planktonic larvae settle out of the water column (Saunders and Metaxas 2007) and grow to form large, encrusting colonies on kelps (Berman et al. 1992; Scheibling et al. 1999). Encrustation by *M. membranacea* weakens the tissue of kelp blades (Krumhansl et al. 2011), increasing their susceptibility to breakage during periods of high wave action (Lambert et al. 1992; Scheibling and Gagnon 2009). Consequently, *M. membranacea* has been linked to large-scale declines in kelp populations in its invasive range (Filbee-Dexter et al. 2016). Since the abundance of the bryozoan increases with temperature, its negative impact on kelp bed ecosystems is predicted to increase with ocean warming due to climate change (Saunders et al. 2010; Denley et al. 2019a). Because of the ecological and economic importance of kelp bed ecosystems (Wernberg et al. 2019), robust SDMs are needed for the abundance of *M. membranacea* in the northwest Atlantic to inform critical management strategies.

Past studies have identified physical environmental factors and biological characteristics of kelp beds affecting the abundance of multiple life stages (newly settled and adult colonies) of the bryozoan in its invasive range. Sea water temperature can regulate abundance, rate of development, and growth of newly settled and adult colonies (Saunders and Metaxas 2007; Scheibling and Gagnon 2009; Caines and Gagnon 2012). Wave exposure may have a negative quadratic relationship with the abundance of newly settled colonies (Caines and Gagnon 2012), potentially by affecting larval delivery and recruitment, and may impact adult abundance through effects on feeding (Pratt 2008; Arkema 2009). Depth has been positively correlated with the abundance of newly settled colonies, possibly due to the influence of non-temperature-related abiotic factors on larval abundance (Saunders and Metaxas 2007; Denley and Metaxas 2017). The species of kelp substrate can influence the abundance of both settlers and adults due to differences in kelp growth and morphology (Saunders and Metaxas 2009a). Lastly, the relative abundance of kelp species within mixed species kelp beds may affect the abundance of adult colonies of *M. membranacea* by altering its population dynamics (Denley et al. 2019a).

Although past research has identified multiple factors affecting the abundance of *M. membranacea*, existing knowledge is insufficient to select the most relevant predictors for use in SDMs for this species. Firstly, there is a mismatch between the large spatial (> 1000 km²; Fig. 1) and temporal

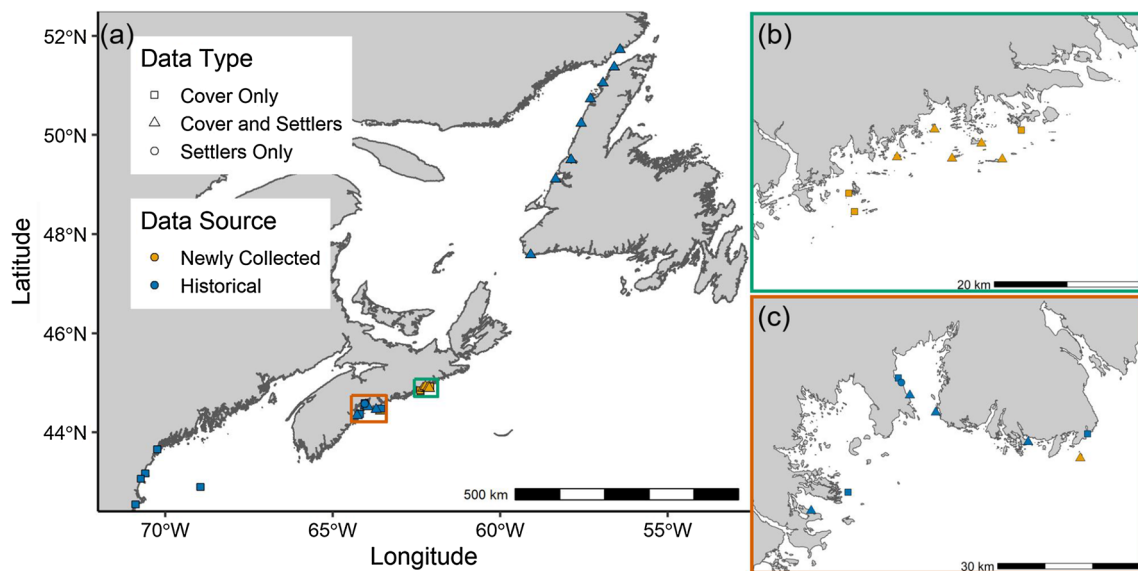


Fig. 1 **a** Locations of historical (1987–2017) and newly collected (2018–2019) observations of percent cover ($n=549$) and settler density ($n=412$) of *Membranipora membranacea* in the northwest

Atlantic Ocean. Insets show sites in **b** the Eastern Shore Islands and **c** the southwestern shore of Nova Scotia, corresponding to green and orange boxes in **a**, respectively

(decadal: 1987–present; Berman et al. 1992) scales of the invasion of *M. membranacea* and those of these past studies, which were conducted over limited temporal ($n=2–3$ years; e.g. Saunders and Metaxas 2009a; but see Scheibling and Gagnon 2009) and geographic (<50 km; e.g. Denley and Metaxas 2017; but see Caines and Gagnon 2012) ranges. Additionally, the relative importance of the predictors of the abundance of *M. membranacea*, as well as the absolute importance (sensu Galipaud et al. 2017) of some candidate predictors (e.g. wave exposure and kelp density), have not been established. Lastly, multiple details regarding the effects of temperature and wave exposure on the abundance of newly settled and adult colonies of the bryozoan are not well-established, including: (1) the temporal windows over which these metrics should be considered; (2) the shapes of the potential relationships between wave exposure and the abundance of *M. membranacea*; and (3) whether the mean or thermal integral (cumulative sum) of temperature over a given temporal window is the superior metric for prediction.

Although SDMs should ideally employ the most ecologically relevant, proximal predictors, data for these are often unavailable at the requisite spatiotemporal scales (Austin 2007; Williams et al. 2012). For example, in the case of benthic marine organisms, in situ temperature at depth (IST) is often an ecologically relevant proximal predictor (e.g. Schiel et al. 2004; Clarke et al. 2009; Hiddink et al. 2015). However, because of the limited availability of IST data, the distal predictor sea surface temperature (SST) is often the only option for large-scale modelling efforts (Stobart et al. 2016). When SST is used in SDMs where observations have been

collected from multiple depths, depth should also be incorporated into the models to account for depth-based temperature variability (Duffy and Chown 2017). However, SST and depth may not be an adequate approximation of IST in all cases and relationships are not consistent in space and time (Castillo and Lima 2010; Stobart et al. 2016). Therefore, when possible, it is important to validate SST as a proxy for IST prior to using it as a predictor in SDMs.

In this study, we demonstrate a detailed, biologically informed process of predictor variable selection that is often absent from the current SDM literature (Mod et al. 2016; Fourcade et al. 2018; Gardner et al. 2019), using the invasive bryozoan *M. membranacea* as a case study. Specifically, we model the relationships between two life-history stages (newly settled and adult colonies) of the bryozoan and their potential predictors using the largest available dataset to date of observations throughout the northwest Atlantic Ocean. After using past findings and ecological reasoning to identify relevant predictor variables and their appropriate metrics, we take an information-theoretic, multi-model inference approach to fulfill the following objectives: (1) determine the optimal metrics and most relevant temporal scales for environmental predictors of abundance; (2) quantify the relationships and determine the importance of environmental predictors in driving the abundance of different life-history stages; (3) compare our quantified relationships and variable importance to the results of past studies conducted at smaller spatiotemporal scales; and (4) compare the performance of models based on a distal (SST) and proximal (IST) predictor of temperature. This study provides a

conceptual framework for selecting variables for inclusion in SDMs that is especially applicable to organisms with numerous potential drivers of abundance for which ample species and predictor data are available.

Materials and methods

Dataset

We compiled a dataset of all available observations of the abundance of adult colonies (measured as the percent cover of adult colonies per unit area of kelp blade) and newly settled colonies (settler density, measured as the number of newly settled colonies per unit area of kelp blade) of *M. membranacea* on the three numerically dominant kelp species (*Agarum clathratum*, *Laminaria digitata*, and *Saccharina latissima*) in the northwest Atlantic Ocean, along with data on their potential environmental drivers (depth, IST, kelp density, kelp substrate [species], SST, and wave exposure). The dataset included historical data from the literature spanning over 2800 km and 30 years ($n = 20$ and 13 sites, for percent cover and settler density, respectively) and new data we collected from 9 additional sites on the Atlantic coast of Nova Scotia, Canada in 2018 and 2019 (Fig. 1, Table S1; for details, see Online Resource 1, S1).

Calculation of response variables

Percent cover of *M. membranacea* in all previous studies and new collections was either measured directly on individual kelp blades or estimated from imagery (as for our newly collected data as detailed in Online Resource 1, S1.1). We found no difference in estimates of percent cover between methods (Online Resource 1, S2) and, therefore, combined percent cover data from all studies irrespective of the collection method.

Although all previous studies and our new collections used the same method to enumerate settler density on individual collected kelp blades (for details see Online Resource 1, S1.1), the definition of a settler varied among studies: settlers were defined either as colonies < 1 cm in diameter or as colonies with ≤ 2 zooid rows (typically 0.5–0.9 mm in diameter; Saunders and Metaxas 2007). We used concurrent observations of abundance for multiple colony size classes from Caines and Gagnon (2012) to generate a relationship between colonies ≤ 2 zooids and < 1 cm in diameter. We then used this relationship to estimate the abundance of colonies with ≤ 2 zooid rows in studies which defined settlers as colonies < 1 cm (Online Resource 1, S3). This allowed us to combine data across studies by standardizing settlers as colonies with ≤ 2 zooid rows.

We averaged percent cover and settler density across all individuals of a given kelp species within each depth, site, and sampling date, and bounded the data between May and November to encompass the approximate annual growth period of *M. membranacea* (Table S1; Online Resource 1, S4).

Identification and calculation of relevant predictor variables

Temperature

We obtained time series of daily SST data from NOAA's "CoralTemp" dataset (NOAA Coral Reef Watch 2020) at a 5-km spatial resolution for the 6 months leading up to the sampling date for each observation of percent cover and settler density in our dataset. In cases where SST data were unavailable at a given site, we used the mean value from all grid cells within a 6-km radius of the site. We also compiled IST data corresponding to observations of percent cover and settler density from past studies (only available from sites in Nova Scotia; Table S1).

We calculated multiple temperature metrics, consisting of the mean and thermal integral over different time periods prior to a given observation of percent cover or settler density, for both IST and SST data (Table 1). For each period prior to the date of observation of percent cover or settler density, we calculated thermal integrals by summing standardized daily temperatures (calculated by adding 1.8 °C to daily average temperatures to avoid negative values; Saunders and Metaxas 2007). To calculate thermal integral, gaps in IST data ($n = 1$ –15 days) were filled using the mean temperature for the given month, site, and depth. Mean temperature metrics were calculated as the mean of daily average temperatures over the given period prior to the date of observation of percent cover or settler density. Thermal integral can effectively predict percent cover (Scheibling and Gagnon 2009; Caines and Gagnon 2012) and settler density of *M. membranacea* (Saunders and Metaxas 2007; Caines and Gagnon 2012). However, we also investigated mean temperature metrics since they do not rely on time series of continuous daily temperature and, therefore, offer greater flexibility in the temperature data that can be used in predictive models. We investigated 3-month temperature metrics as predictors of percent cover based on the findings of Scheibling and Gagnon (2009) that the 3-month integral of temperature had a strong relationship with annual peak percent cover. We also assessed 6-month temperature metrics to account for the potential effects of temperature on *M. membranacea* over longer timescales that encompass multiple stages in its lifecycle (larvae, reproduction, etc.; Saunders and Metaxas 2008; Caines and Gagnon 2012). For settler density, we investigated 1-month temperature

Table 1 Summary of candidate temperature (SST or IST) and REI metrics considered as predictors of percent cover and settler density

Model type	Temperature		REI	
	Metric	Abbreviation	Metric	Abbreviation
Settler density models only	3-month integral	1-I	1-month linear	1-Lin
	1-month mean	1-M	3-month logarithmic	1-Log
			1-month quadratic	1-Q
Percent cover and settler density models	3-month integral	3-I	3-month linear	3-Lin
			3-month logarithmic	3-Log
	3-month mean	3-M	3-month quadratic	3-Q
			6-month linear	6-Lin
	6-month integral	6-I	6-month logarithmic	6-Log
			6-month quadratic	6-Q
	6-month mean	6-M	Calendar-year linear	Y-Lin
			Calendar-year logarithmic	Y-Log
			Calendar-year quadratic	Y-Q

Time periods in the metrics refer to the period prior to a given observation of percent cover or settler density over which temperature or wind data (for REI) were averaged or summed. One-month metrics were only considered as candidate predictors in models for settler density

metrics since larvae of *M. membranacea* have an average planktonic larval duration of 4 weeks (Yoshioka 1982), and 3- and 6-month metrics because the abundance of settlers may be explained by thermal indices which affect both the planktonic larval period and the adult colonies producing the larvae (Caines and Gagnon 2012).

We examined the relative effects of seasonal temperature variation (day of the year) and site- or year-specific thermal anomalies using residual regression (Graham 2003). We generated generalized additive models (GAMs) of the non-linear relationship between day of the year (1–365; independent variable) and all observations of each temperature metric (log-transformed if necessary to satisfy model assumptions) and used the residuals to represent temperature anomalies in subsequent GLMMs (SST anomaly in [Methods: Importance of seasonal SST and SST anomaly in predicting percent cover and settler density](#)). We used the day of the year to represent seasonal average temperature variation in the northwest Atlantic (seasonal SST in [Methods: Importance of seasonal SST and SST anomaly in predicting percent cover and settler density](#)).

Wave exposure

We calculated wave exposure using the relative wave exposure index (REI) after Krumhansl and Scheibling (2011):

$$REI = \sum_{i=1}^{16} (V_i \times W_i \times F_i) \quad (1)$$

where V_i and W_i are average wind speed (km h^{-1}) and wind frequency over a given temporal window, and F_i is fetch

(km; bounded at 2000 km for a given direction) from the i th cardinal direction (north, north-northeast, northeast, east-northeast, etc.), in 16 bins of 22.5° centred on each direction. We obtained wind data from weather stations within a maximum distance of ~ 100 km from each site, based on data availability, proximity, and geographical representativeness, from Environment and Natural Resources Canada's Historical Climate Database (<http://climate.weather.gc.ca>) for Canadian sites and NOAA's National Data Buoy Center (<https://www.ndbc.noaa.gov>) for sites in the Gulf of Maine (Fig. S3). REI was linearly related to in situ wave exposure measured as bottom orbital velocity at 5 sites (Online Resource 1, S5).

To calculate metrics of REI for each observation, we first averaged hourly wind data over the same time periods as for the temperature metrics (1, 3 and 6 months), as well as over the calendar year of a given observation to describe the average exposure conditions at that site. We then transformed the time-averaged REI values into linear, logarithmic, and quadratic forms to test different hypotheses regarding relationships between REI and percent cover and settler abundance of *M. membranacea* (Table 1). A negative linear relationship between REI and the abundance of *M. membranacea* may be possible because high wave exposure could physically inhibit larval settlement (Koehl 2007) and slow the growth of adult colonies due to the energetic costs of producing protective structures in response to high wave action (Bayer et al. 1997). Similarly, a negative log-linear relationship could occur if a threshold exists beyond which settlement or colony growth is impeded by the above mechanisms. Lastly, a quadratic relationship reflects possible inhibitory effects of both high

and low wave exposure on settlement (Pawlik and Butman 1993) and the feeding success of adult colonies (Pratt 2008; Arkema 2009), respectively.

Kelp density

For sites where all three kelp species were present, we calculated the site- and depth-specific mean density of each species of kelp from July–September (summer), October–December (autumn), January–March (winter), and April–June (spring). We then calculated the relative densities of *L. digitata* and *S. latissima* (species-specific kelp density divided by total kelp density; $Kelp_R$) and total kelp density (kelp density summed across all 3 species; $Kelp_T$) for each depth, site, and period. We considered quadratic $Kelp_R$ (of *L. digitata* and *S. latissima*) as predictors of percent cover of *M. membranacea* based on the prediction of Denley et al. (2019a) that percent cover is highest for mixed-species kelp beds (i.e. percent cover should be greatest at intermediate values of $Kelp_R$). We hypothesised that, if larval supply limits settlement of *M. membranacea* as for other marine invertebrates (e.g. Gaines et al. 1985), a linear relationship could exist between $Kelp_T$ and settler density: positive if physical factors such as the roughness (Eckman 1990) and internal flow attenuation (Eckman et al. 1989) of kelp beds enhance settlement, or negative if the presence of kelp does not induce settlement (Denley and Metaxas 2017) but higher densities of kelp spread settlers over a larger surface area of kelp and thus decrease their density. Although *M. membranacea* occurs primarily on kelp substrates (Yorke and Metaxas 2012) and its probability of presence or absence may therefore be affected by the presence of kelp, it should be noted that all observations of cover and settler density in our dataset were from mixed-species kelp beds, and our goal was to test relative and total kelp density as predictors of abundance, not presence or absence, of the bryozoan.

Preliminary analyses revealed moderate collinearity between the optimal REI metric (log of 6-month REI; see [Methods: Selection of optimal SST and REI metrics for predicting percent cover and settler density](#) and [Results: Optimal SST and REI metrics for predicting percent cover and settler density](#)) and the relative density of *L. digitata* ($r=0.43$) and *S. latissima* ($r=-0.34$). To examine the effect of kelp density on percent cover independently from that of REI, we conducted residual regression (Graham 2003) using linear models with $Kelp_R$ as the dependent variable and the log of 6-month REI as the independent variable. We used the residuals of these models as the $Kelp_R$ variables in subsequent GLMMs for percent cover.

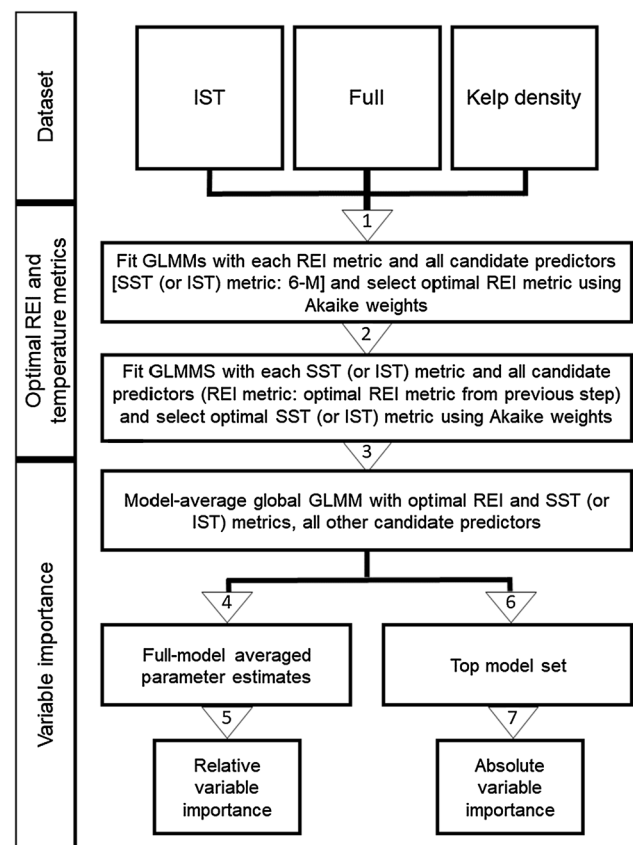


Fig. 2 Flow of data processing and modelling procedures used to assess the importance of predictors in contributing to observed patterns in percent cover and settler density of *M. membranacea* in the northwest Atlantic Ocean over broad spatial and temporal scales. Numbered arrows correspond to modelling procedures described in [Methods: Selection of optimal SST and REI metrics for predicting percent cover and settler density](#) and [Methods: Importance of candidate predictors of percent cover and settler density](#)

Assessment of predictor variables

We used an information-theoretic multi-model inference approach, based on generalized linear mixed models (GLMMs), to assess the relative and absolute importance of candidate predictors (and metrics) and the relationships of these predictors with percent cover and settler density of *M. membranacea* (Fig. 2; for complete detailed schematic, see Fig. S5). We conducted all analyses in R, version 3.6.2 (R Core Team 2020) in the R Studio environment (RStudio Team 2020). We fit GLMMs using glmmTMB (Brooks et al. 2017), validated them with DHARMa (Hartig 2021), and conducted model averaging with MuMIn (Barton 2019). See Online Resource 1, S7 for a full list of packages used.

Selection of optimal SST and REI metrics for predicting percent cover and settler density

To determine the optimal REI and SST metrics for predicting percent cover and settler density, we compared the performance of GLMMs containing each candidate metric (Table 1) using ratios of their Akaike weights (Wagenmakers and Farrell 2004). First, we standardized all continuous predictor variables by subtracting the sample mean from each observation (Schielzeth 2010) and dividing by 2 standard deviations (Gelman 2008), and applied a sum-to-zero contrast to the kelp substrate variable (Schielzeth 2010). Standardization allowed comparison of parameter estimates among models to assess relative effect size, while the sum-to-zero contrast simplified the interpretation of model-averaged interaction terms (Schielzeth 2010). After standardization, we examined collinearity between predictors by calculating pairwise correlations and variance inflation factors (Dormann et al. 2013). We then fit GLMMs composed of: (1) all candidate predictor variables except kelp density (depth, kelp substrate, REI, and SST) as fixed effects; (2) an interaction between temperature and kelp substrate, which was identified as significant in preliminary analyses; and (3) crossed random intercepts for study site and study year, to account for the blocked nature of the data within these factors (Table 2). GLMMs were fit using maximum likelihood. For models of percent cover, we used a beta model with an intercept-only zero-inflation term and a logit link function (Ospina and Ferrari 2012). For models of settler density, we

used a negative binomial model with a log link function. We validated the GLMMs by visually assessing QQ plots and plots of simulated residuals vs. fitted values.

To determine optimal metrics, we first compared models fit with each REI metric, using the same temperature metric (6-month mean of SST [SST 6-M]) as the SST metric in each model (Fig. 2, arrow 1). We selected the optimal REI metric by identifying the metric generating the model with the highest Akaike weight, and compared its performance relative to other REI metrics using ratios of their Akaike weights (evidence ratios; Wagenmakers and Farrell 2004). Using this optimal REI metric in each model, we then compared models fit using each SST metric, and selected the optimal SST metric using Akaike weights (Fig. 2, arrow 2).

Importance of candidate predictors of percent cover and settler density

For percent cover and settler density, we fit a global GLMM (using the same procedure described above in [Methods: Selection of optimal SST and REI metrics for predicting percent cover and settler density](#)) with depth, kelp substrate, the optimal REI metric, the optimal SST metric, an interaction between SST and kelp substrate, and random intercepts for site and year. To evaluate variable importance, we conducted model averaging on the global GLMM and derived a complete model set, consisting of all possible sub-models (combinations of fixed effects) of the global GLMM (Fig. 2, arrow 3; Burnham and Anderson 2002). We calculated

Table 2 General forms of global generalized linear mixed models for percent cover and settler abundance, where Percent Cover_{ijk} and Settler Density_{ijk} are the *k*th observation at site *i* in year *j*, μ is the mean,

α is the probability of observing 0, θ is unknown parameter controlling the variance, and *k* is the dispersion parameter

Equation	Percent cover	Settler density
Error distribution	Percent Cover _{ijk} \sim BEZI($\alpha, \mu_{ijk}, \theta$)	Settler Density _{ijk} \sim NB(μ_{ijk}, k)
Conditional mean function	$E(\text{Percent Cover}_{ijk}) = (1 - \alpha)\mu_{ijk}$	$E(\text{Settler Density}_{ijk}) = \mu_{ijk}$
Variance function	$\text{var}(\text{Percent Cover}_{ijk}) = \alpha(1 - \alpha)(-\mu_{ijk})^2 \times (1 - \alpha) \frac{\mu_{ijk}(1 - \mu_{ijk})}{1 + \theta}$	$\text{var}(\text{Settler Density}_{ijk}) = \mu_{ijk} + \frac{\mu_{ijk}^2}{k}$
Model equation	$\text{logit}(\mu_{ijk}) = \text{Depth}_{ijk} + \text{Kelp Substrate}_{ijk} + \text{REI}_{ijk} + \text{Temperature}_{ijk} + (\text{Kelp}_{\text{LD}_{ijk}} + \text{Kelp}_{\text{SL}_{ijk}}) + \text{Temperature}_{ijk} \times \text{Kelp Substrate}_{ijk} + \text{Site}_i + \text{Year}_j$	$\log(\mu_{ijk}) = \text{Depth}_{ijk} + \text{Kelp Substrate}_{ijk} + \text{REI}_{ijk} + \text{Temperature}_{ijk} + (\text{Kelp}_{\text{T}_{ijk}}) + \text{Temperature}_{ijk} \times \text{Kelp Substrate}_{ijk} + \text{Site}_i + \text{Year}_j$
Random intercept distributions	$\text{Site}_i \sim N(0, \sigma_{\text{Site}}^2)$ $\text{Year}_j \sim N(0, \sigma_{\text{Year}}^2)$	$\text{Site}_i \sim N(0, \sigma_{\text{Site}}^2)$ $\text{Year}_j \sim N(0, \sigma_{\text{Year}}^2)$

BEZI represents a zero-inflated beta distribution and NB represents a negative binomial distribution. LD and SL are abbreviations for *L. digitata* and *S. latissima*, respectively

full-model averaged parameter estimates (Symonds and Moussalli 2011) and their unconditional 95% confidence intervals (Burnham and Anderson 2004) for each predictor variable from the complete model set (Fig. 2, arrow 4). We assessed the relationship between the predictor variable and percent cover or settler density based on the sign of its estimate (i.e. positive or negative; or the relative magnitude of the parameter estimates for each species, in the case of kelp substrate) and compared the magnitude of parameter estimates between predictors to assess relative variable importance (Fig. 2, arrow 5).

To determine absolute variable importance, we ranked the models in the complete model set in order of increasing values of Akaike's Information Criterion (AIC) and derived a top model set containing models for which ΔAIC (the increase in AIC of a given model compared to the top-ranked model) was less than 7, creating a top model set 95% likely to contain the actual best model (Fig. 2, arrow 6; Burnham and Anderson 2002; Richards 2005). We excluded models for which a less complex nested model with an equal or lower AIC value was also included in the top model set (Richards 2008; Richards et al. 2011). We considered predictor variables retained in all models within the top model set to be informative predictors (*sensu* Leroux 2019) of percent cover or settler density (Fig. 2, arrow 7). If a predictor variable was not included in all models in the top model set, we used ΔAIC to compare models with and without the parameter and also examined if the 95% confidence interval for its parameter estimate included zero to further assess its importance (Grueber et al. 2011). To test for significant differences among kelp substrate species and kelp-SST interactions, we conducted Tukey tests using the top-ranked model in each model set. Lastly, we determined the goodness-of-fit of the models in the top model set using Nakagawa's R^2 (Nakagawa et al. 2017).

We investigated the relative and absolute importance of kelp density separately from the other predictors (Online Resource 1, S6), since substantially fewer data on kelp density were available and therefore assessing the importance of kelp density with the other predictors would have greatly reduced our sample size. However, kelp density was not an informative predictor for explaining patterns of percent cover and settler density (Online Resource 1, S6). Therefore, we excluded kelp density and constructed models based on the full available dataset for all other analyses.

Importance of seasonal SST and SST anomaly in predicting percent cover and settler density

Our dataset contained observations of percent cover and settler density over the entire growing season of *M. membranacea* (May–November) from different sites over different years and thus may have been influenced by both seasonal

patterns in temperature and temperature variability between sites and years. To investigate the relative influence of these components of temperature variability as predictors of the abundance of *M. membranacea*, we constructed global models for cover and settler density with depth, kelp substrate, the optimal REI metric, seasonal SST and SST anomaly (using the optimal SST metric) as fixed effects, an interaction between kelp substrate and seasonal SST, and site and year as random intercepts. We then derived complete model sets from these global models, and assessed variable importance using model averaging.

IST- versus SST-based models

To investigate the suitability of SST as a proxy for IST when predicting the abundance of *M. membranacea*, we examined differences in relative and absolute variable importance between models based on temperature data from IST and SST. For this comparison, we ran models using the subset of the full datasets for which IST data were available (IST datasets; $n = 165$ for percent cover, $n = 192$ for settler density; Table S1). First, we identified the optimal REI and temperature metrics using IST- and SST-based models for percent cover and settler density, using the procedure in [Methods: Selection of optimal SST and REI metrics for predicting percent cover and settler density](#). Then, we assessed relative variable importance and absolute variable importance as described in [Methods: Importance of candidate predictors of percent cover and settler density](#). Finally, we compared variable importance results and top model fit from the IST- and SST-based models for percent cover and settler density.

Results

Optimal SST and REI metrics for predicting percent cover and settler density

The optimal SST metric for explaining observed patterns in percent cover was the 6-month mean of SST (SST 6-M); however, the Akaike weight of the model containing SST 6-M was only 1.11 times greater than that containing the 6-month integral of SST (SST 6-I; Table 3a). Akaike weights of the models with both 6-month metrics were approximately three orders of magnitude greater than those with both 3-month metrics. For predicting settler density, the optimal SST metric was SST 6-I (Table 3b), with Akaike weight of the model containing SST 6-I being 1.80 times greater than that with SST 6-M. Akaike weights for models with both 6-month metrics were over three orders of magnitude greater than those for 3-month and 1-month metrics.

The log of 6-month REI (REI 6-Log) was the optimal REI metric for predicting percent cover (Table 3a), followed closely by the log of 3-month REI (REI 3-Log) and the log of calendar-year REI (REI Y-Log). REI Y-Log was the optimal REI metric for predicting settler density, with the Akaike weight of the model containing REI Y-Log being 1.06 times greater than that with REI 6-log (Table 3b).

Importance of candidate predictors of percent cover and settler density

Our top-ranked GLMMs for percent cover and settler density both attained marginal R^2 values of over 0.5 (Table 4a, b). SST was the most important predictor of both percent cover and settler density, with a parameter estimate more than twice as large as the next largest (kelp substrate; Fig. 3a, b). Both seasonal SST and SST anomaly were informative individual predictors of percent cover and settler density of *M. membranacea*, both appearing in all models in the top model sets; however, seasonal SST was approximately two times more important than SST anomaly in both cases (Fig. S6, Table S2).

SST had a positive relationship with cover and settler density but the strength of this relationship was strongly dependent on kelp substrate (Fig. 4d, h). The positive relationship between SST and percent cover was stronger for *L. digitata* than for *S. latissima* or *A. clathratum* while the positive relationship between SST and settler density was significantly lower for *A. clathratum* than *L. digitata* or *S. latissima* (Tukey test, $p < 0.01$; Fig. 3a, b). The second most important predictor of percent cover and settler density was kelp substrate (Fig. 3a, b). Mean percent cover was significantly higher on *L. digitata* than on *S. latissima* or *A. clathratum* while mean settler density was significantly lower on *A. clathratum* than on *L. digitata* or *S. latissima* (Tukey test, $p < 0.01$; Fig. 3a, b, 4b, f). Depth was the third most important predictor of percent cover and the least important for settler density (Fig. 3a, b) with a negative and positive relationship, respectively (Fig. 4a, e). The negative, logarithmic relationship with REI was the least important predictor of percent cover (Fig. 3a, 4c) although it was the third most important predictor of settler density (Fig. 3b, 4g). For percent cover, all candidate variables except REI were included in both models in the top model set; there was a ΔAIC of < 1 when REI was removed from the top model in the model set (Table 4a) and its 95% confidence interval included zero (Fig. 3a), providing little evidence that it is an informative predictor. For settler density, all predictors appeared in both models in the top model set except REI, which only appeared in the top-ranked model; there was a ΔAIC of > 4 when REI

was removed in the second-ranked model (Table 4b) and its 95% confidence interval did not include zero (Fig. 4b), indicating that it is an informative predictor.

IST- vs SST-based models

We compared the results of SST-based models to those of IST-based models to assess the suitability of SST as a proxy for IST for predicting the abundance of *M. membranacea* (for detailed results, see Tables S4 and S5). The optimal temperature metric was the 3-month integral in SST-based models and the 3-month mean in IST-based models of percent cover (Tables 5a; S4a, b). For settler density, the optimal temperature metric was the 3-month integral in SST-based models and the 3-month integral in IST-based models (Table 5b). For both percent cover and settler density, there were no notable differences in the rankings of REI metrics between IST- and SST-based models (Table S4). For IST- and SST-based models of settler density, quadratic 6-month mean of REI (REI 6-Q) produced the model with the highest Akaike weight. However, the quadratic relationship between REI 6-Q and settler density seems to be an artifact of the dataset (few data-points at intermediate values; Fig. S7) rather than a real effect. Consequently, we elected to use the log of 1-month mean REI (REI 1-Log), which had a parameter estimate of almost identical magnitude to REI 6-Q, as the optimal REI metric for subsequent analyses.

For percent cover, relative variable importance rankings were similar for IST- and SST-based models except depth, which had a higher relative importance in SST-based models than in IST-based models (Table 5a, Fig. S9a, c). Absolute variable importance was also the same in IST- and SST-based models except depth, which was an informative predictor in SST- but not IST-based models (Tables 5a, S5a, b, Fig. S9a, c). The top-ranked IST-based model performed better in terms of model fit (higher marginal R^2 and lower AIC) than the top-ranked SST-based model (Tables 5a, S5a, b). For settler density, the relative importance of depth and REI was higher in IST-based models than in SST-based models, while the rankings of temperature, kelp substrate and REI were the same (Table 5b, Fig. S9b, d). Absolute variable importance did not differ between IST- and SST-based models, with depth, temperature and kelp substrate being informative. The top-ranked IST-based model had lower AIC but also slightly lower marginal R^2 than the top-ranked SST-based model (Tables 5b, S5c, d).

Table 3 Results of GLMMs comparing REI and SST metrics as predictors of a) percent cover and b) settler density

(a)				
Variable	Metric	AIC	Δ AIC	weight
REI	6-Log	– 1282.436	0.000	0.258
	3-Log	– 1282.130	0.306	0.222
	Y-Log	– 1281.523	0.913	0.164
	6-Lin	– 1280.866	1.570	0.118
	Y-Lin	– 1279.936	2.500	0.074
	3-Lin	– 1279.766	2.670	0.068
	6-Q	– 1278.873	3.563	0.044
	Y-Q	– 1277.938	4.498	0.027
	3-Q	– 1277.780	4.656	0.025
SST	6-M	– 1282.436	0.000	0.527
	6-I	– 1282.222	0.214	0.473
	3-I	– 1265.605	16.831	<0.001
	3-M	– 1265.438	16.998	<0.001
(b)				
Variable	Metric	AIC	Δ AIC	weight
REI	Y-Log	3699.284	0.000	0.339
	6-Log	3699.392	0.108	0.321
	6-Q	3700.559	1.275	0.179
	Y-Q	3702.862	3.578	0.057
	3-Log	3704.073	4.789	0.031
	1-Log	3705.668	6.384	0.014
	1-Lin	3705.900	6.616	0.012
	3-Lin	3706.007	6.723	0.012
	Y-Lin	3706.042	6.758	0.012
	6-Lin	3706.210	6.926	0.011
	3-Q	3707.164	7.880	0.007
	1-Q	3707.329	8.045	0.006
	6-I	3698.044	0.000	0.650
SST	6-M	3699.284	1.240	0.350
	3-I	3722.245	24.201	<0.001
	3-M	3722.513	24.469	<0.001
	1-M	3834.331	136.287	<0.001
	1-I	3834.575	136.530	<0.001

Models of the form shown in Table 2 (excluding kelp density as a predictor) were fit using each REI and SST metric. For comparison of REI metrics, 6-month mean of SST (SST 6-M) was the temperature metric in the model. For comparison of SST metrics, the optimal REI metric was employed as the REI metric in the model. AIC, Δ AIC (change in AIC relative to top-ranked model) and Akaike weight were calculated for each model. REI and SST metrics are ranked in order of descending model performance and are abbreviated as per Table 1

Discussion

Following a detailed process of variable selection, the top-ranked models in this study explained a substantial amount of the variability (> 50%; Cohen 1988; Møller and Jennions 2002) in the abundance of multiple life stages of the invasive bryozoan *M. membranacea*. With few exceptions, variable importance and the relationships between environmental predictors and abundance revealed in our large-scale study

aligned with those of past studies conducted at smaller spatiotemporal scales. This case study using *M. membranacea*, a well-studied invasive species for which abundance data were available, may represent a best-case scenario for variable selection for SDMs. However, the framework we used can be applied widely in species distribution modelling to select relevant predictor variables, including for organisms for which data are available as presence-absence or presence-only (Fig. 5).

Table 4 Sets of top models (models with $\Delta\text{AIC} < 7$) explaining observed patterns in (a) percent cover and (b) settler density of *M. membranacea* in the northwest Atlantic. REI and SST model terms are abbreviated as per Table 1

(a)								
Model	df	logLik	AIC	ΔAIC	Akaike Weight	Marg R^2	Cond R^2	Random effect SD
Depth + REI 6-Log + Kelp Substrate + SST 6-M + Kelp Substrate:SST 6-M	12	653.218	– 1282.436	0.000	0.581	0.588	0.766	Site = 0.566 Year = 0.142
Depth + Kelp Substrate + SST 6-M + Kelp Substrate:SST 6-M	11	651.858	– 1281.717	0.719	0.405	0.558	0.758	Site = 0.592 Year = 0.142
(b)								
Model	df	logLik	AIC	ΔAIC	Akaike Weight	Marg R^2	Cond R^2	Random effect SD
Depth + REI Y-Log + Kelp Substrate + SST 6-I + Kelp Substrate:SST 6-I	11	– 1838.022	3698.044	0.000	0.891	0.542	0.778	Year = 0.845 Site = 0.654
Depth + Kelp Substrate + SST 6-I + Kelp Substrate:SST 6-I	10	– 1841.444	3702.888	4.844	0.079	0.494	0.775	Year = 0.849 Site = 0.788

Marg. and Cond. R^2 are the marginal and conditional R^2 values for the GLMM. Random Effect SD is the standard deviation of each random intercept

Our comparisons of models based on different temperature data sources indicated that the suitability of distal predictors (e.g. depth and SST) as proxies for their in situ counterparts (e.g. IST) should be investigated whenever possible in the context of SDM (Fig. 5, step 3). Rankings of metrics and variable importance were similar and model fit was only slightly improved in IST- compared to SST-based models of percent cover, indicating that depth and SST can be used to approximate IST in SDMs for percent cover of *M. membranacea*. However, rankings of temperature metrics and the relative importance of depth differed between SST- and IST-based models of settler density, potentially due to confounding effects of another depth-dependent variable on settler density. The differences between SST- and IST-based models for settler density, as well as the spatial limitation of the IST dataset to a subset of our study region (Nova Scotia), underscore the necessity to evaluate the suitability of SST and depth as a proxy for IST on a case-by-case basis.

Our results highlight the importance of carefully selecting metrics used to summarise environmental data employed in SDMs to match the scale of the ecological processes affecting the organism or life stage being modelled (Fig. 5, step 4). First, the 6-month thermal windows of the optimal SST

metrics in this study were consistent with previous studies, showing that the abundance of both settlers and adult colonies depends on the thermal history affecting the current and previous life stage (Saunders et al. 2010; Caines and Gagnon 2012). Second, we found that thermal integrals of SST explained patterns in settler density better than mean temperature, but that thermal integrals and means explained patterns in percent cover equally well. These results were likely caused by differences in thermal sensitivity between settlers and adult colonies of the bryozoan. Early life stages of marine ectotherms are often more sensitive to temperature than adults (Pörtner et al. 2007; Putnam et al. 2010; Pineda et al. 2012; Reglero et al. 2014; Pandori and Sorte 2019) and, as a result, thermal integrals may be superior to mean temperature in predicting the recruitment of *M. membranacea* (Neuheimer and Taggart 2007; Metaxas and Saunders 2009). The calculation of thermal integrals requires daily temperature data whereas means can be calculated from data at a lower temporal resolution. Thus, the suitable temporal resolution of environmental data may differ between organisms or life stages (Fig. 5, step 3).

Additionally, our findings demonstrate that the knowledge gained by assessing relative variable importance in an SDM can affect subsequent decisions on model projection

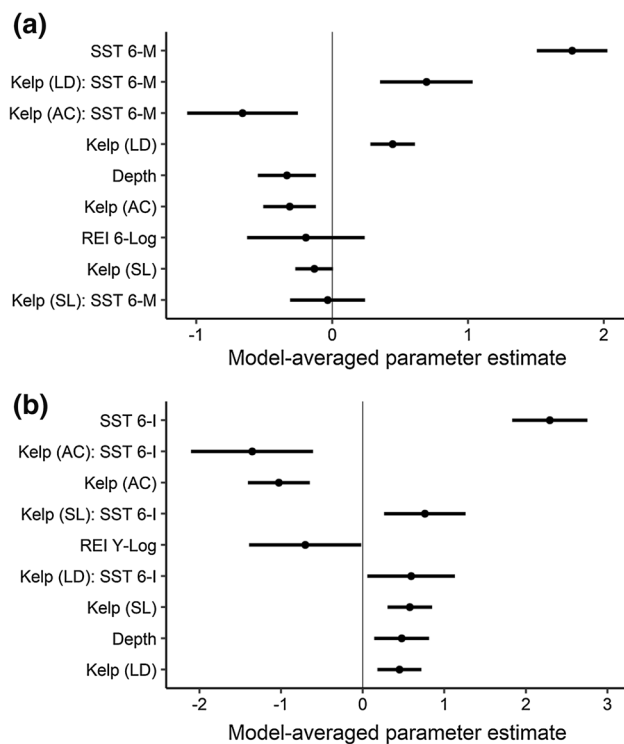


Fig. 3 Full-model averaged standardized parameter estimates for depth, kelp substrate, REI and SST as predictors of **a** percent cover and **b** settler density, arranged in descending order of magnitude. REI and SST metrics are abbreviated as per Table 1. Error bars represent 95% confidence intervals based on unconditional standard error. AC = *A. clathratum*, LD = *L. digitata*, SL = *S. latissima*

and implementation (Fig. 5, step 5). Consistent with previous smaller-scale studies of the bryozoan (Saunders and Metaxas 2007; Scheibling and Gagnon 2009; Caines and Gagnon 2012), temperature was the most important predictor of percent cover and settler density in our models, with a positive effect on abundance in both life stages. Notably, we found that although seasonal changes in temperature were most strongly associated with abundance, temperature anomalies among sites and years were also informative predictors of percent cover and settler density, providing evidence that spatial and interannual difference in temperature drive variability in the abundance of *M. membranacea*. The strong, positive, and direct effect of temperature on abundance reinforces the conclusions of previous studies (Saunders et al. 2010; Denley et al. 2019a) that the abundance of *M. membranacea* will increase with climate change-induced warming in the northwest Atlantic, and emphasises the need for SDMs predicting changes in distribution and abundance under future climate scenarios. Further, kelp substrate was the second most important predictor of both percent cover and settler abundance, suggesting that SDMs predicting future temperature-induced changes in the abundance of

the bryozoan must be projected separately for different kelp substrates.

Even when a variable is an informative predictor in an SDM, the data used to calculate the predictor may not be sufficient to fully capture the ecological mechanism by which the predictor affects the focal organism (Fig. 5, step 6). This is illustrated in this study by the emergence of mean REI over the calendar year as the optimal REI metric for predicting settler density. The negative logarithmic relationship between REI and settler density is ecologically justifiable, implying that settlement sharply decreases beyond a threshold value of REI. Since post-settlement mortality is low for *M. membranacea* (Yoshioka 1982; Denley and Metaxas 2016), this decrease is likely related to an inhibitory effect of wave action on larval settlement (Koehl 2007) rather than an effect on post-settlement mortality. However, it is not clear that calendar year as the optimal temporal window for calculating REI has an ecological basis because it incorporates at least some wind data that could not have affected the bryozoan. Rather, this indicates that our models could not elucidate the real temporal window over which REI has the greatest effect on the settlement of *M. membranacea*. The type of data used to represent an environmental predictor can affect the ability of SDMs to account for and model the details of species-environment relationships, affecting their ecological realism (Fig. 5, step 6). This outcome underscores the importance of carefully assessing and selecting predictor data prior to modelling (Fig. 5, step 3).

Our investigations of kelp density as a candidate predictor of bryozoan abundance illustrate that, in some instances, it can be difficult to determine whether the low importance of a predictor indicates that it is truly uninformative or that the result is an artifact of data analysis. Here, we hypothesised that the total density of kelp would influence settler density and that the relative density of the three dominant kelp species in the northwest Atlantic would affect percent cover of the bryozoan. Contrary to our expectations, we found that neither the density nor species composition of kelps affected the abundance of *M. membranacea*. However, unlike previous studies, our analyses relating relative kelp density to percent cover only focussed on mixed kelp beds where all three kelp species were present; previous predictive models compared population growth of *M. membranacea* for mixed kelp beds and mono-specific beds of the same density (Denley et al. 2019a). Additionally, our datasets for kelp density were considerably smaller than those for the other candidate predictors, possibly affecting our ability to detect an effect of kelp density. In our models, we were able to explain a substantial amount of variance in the abundance of *M. membranacea* despite kelp density being an uninformative predictor. However, in cases where a predictor hypothesised as important is found to be uninformative, it may be necessary

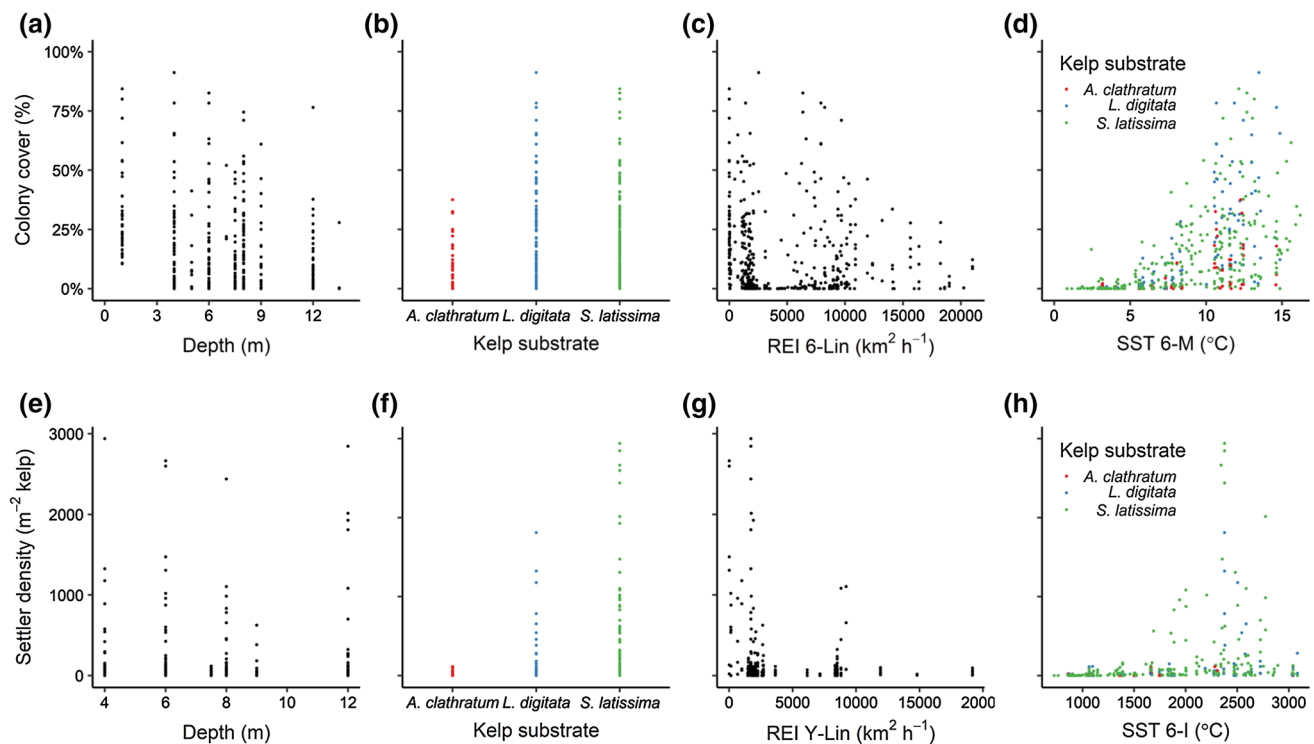


Fig. 4 Candidate predictors of the percent cover **a–d** and settler density **e–h** of *M. membranacea* in the northwest Atlantic. Percent cover of colonies of *M. membranacea* on kelp blades (kelp species-specific means for each depth, site, and sampling date; $n=549$) as a function of **a** depth, **b** kelp substrate species, **c** 6-month mean of REI, and **d**

6-month mean of SST separated by kelp substrate. Settler density (no. individuals per m^2 kelp; kelp species-specific means for each depth, site, and sampling date; $n=412$) as a function of **e** depth, **f** kelp substrate species, **g** mean REI over the calendar year, and **h** 6-month integral SST separated by kelp substrate

Table 5 Comparison of SST- and IST-based models fit using the subset IST dataset for (a) percent cover and (b) settler abundance

(a)						
Temperature Data source	Optimal temperature Metric	Optimal REI Metric	Relative variable Importance	Absolute variable Importance	Top model AIC	Top model R^2
SST	SST 3-I	REI 6-Log	SST > Depth > Kelp Substrate > REI	Depth, Kelp Substrate, SST	− 470.658	0.586
IST	IST 3-M	REI 3-Lin	IST > Kelp Substrate > Depth > REI	Kelp Substrate, IST	− 483.565	0.616
(b)						
Temperature Data source	Optimal temperature Metric	Optimal REI Metric	Relative variable Importance	Absolute variable Importance	Top model AIC	Top model R^2
SST	SST 3-I	REI 1-Log	SST > Kelp > Depth > REI	Depth, Kelp Substrate, SST	1875.898	0.327
IST	IST 1-I	REI 1-Log	IST > Depth > Kelp Substrate > REI	Depth, Kelp Substrate, IST	1857.641	0.275

Optimal temperature and REI metrics indicate the metrics producing models with the largest Akaike weight when compared to models fit using other metrics and are abbreviated as per Table 1. Relative variable importance indicates the ranking of candidate predictor variables by their full-model averaged parameter estimates. Absolute variable importance indicates the variables which were included as predictors in models in the top model set (i.e. informative predictors). Top model R^2 is the marginal R^2 for the top-ranked GLMM in the model set

to deem the resultant model unsatisfactory and return to the beginning of the variable selection process (Fig. 5, step 7).

In this study, we took an a priori variable selection approach and only considered variables that have been identified by previous studies as potential predictors the

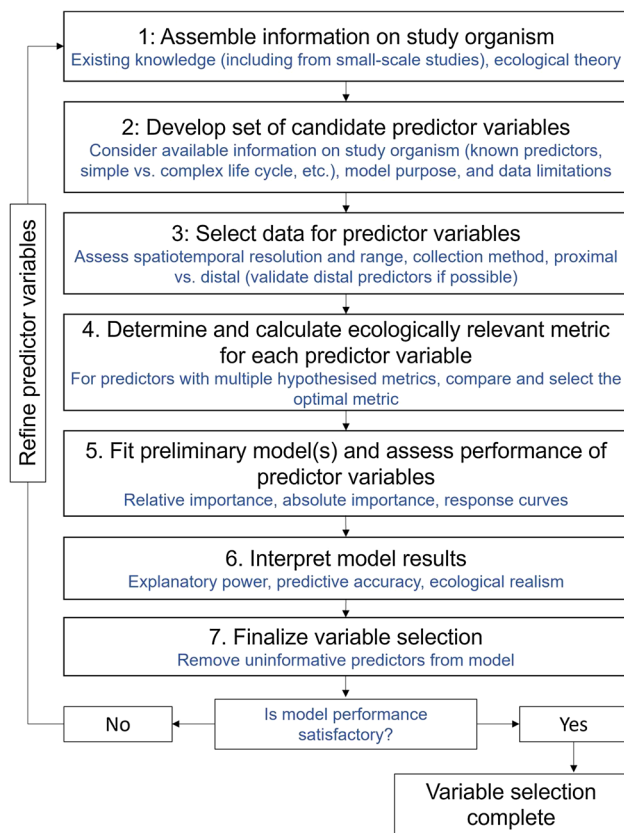


Fig. 5 Conceptual framework for variable selection in species distribution modelling. Numbered black text shows each step in the variable selection process, blue text details considerations at a given step

abundance of *M. membranacea* in the northwest Atlantic Ocean. However, we chose not to include some potential predictors of *M. membranacea* for data-related reasons. Although current velocity may affect settlement (Koehl 2007) and growth (Pratt 2008; Arkema 2009), and both larvae (Stricker et al. 1988) and adult colonies (Saunders and Metaxas 2009b) are planktotrophic, we did not include current velocity or variables related to phytoplankton biomass as predictors in our study because available environmental layers for these predictors did not coincide temporally with our data or were only available at coarser spatial resolution. Observations in our dataset spanned a narrow range of salinity (~30–32 PSU) and preliminary analyses indicated that salinity had no effect on the percent cover or settler density of the bryozoan for our data. The ability of our models to explain a large amount of variability in both percent cover and settlement indicates that these variables are unlikely to exert a large influence on either life stage of the bryozoan within the northwest Atlantic Ocean.

To balance model explanatory power with ecological interpretability and extrapolation ability (Peterson et al. 2011; Petitpierre et al. 2017; Brodie et al. 2020), we chose not to investigate previously unexplored predictors of *M.*

membranacea in this study. The population dynamics of *M. membranacea* have been studied for over 3 decades in its invaded range, and the most important drivers in the region are, therefore, likely to have already been identified. However, studies attempting to fit SDMs for organisms without previously defined predictors, or those aiming to maximize predictive power when interpolating within the same spatiotemporal context of SDM calibration, should consider previously unexplored predictor variables during the variable selection process (Fig. 5, step 2; Petitpierre et al. 2017). Such predictors may include environmental variables, as employed in this study, but also anthropogenic variables, which are less commonly used in SDMs but can be important predictors of distribution, particularly for invasive species (Gallardo et al. 2015; Blanco et al. 2021).

Conclusions

Using a detailed variable selection procedure grounded in ecological reasoning and decades of past research, we determined relevant predictors and their optimal metrics for use in SDMs for multiple life stages of *M. membranacea* in the northwest Atlantic. Our analyses revealed general conclusions relevant to variable selection for SDMs: (1) Relative variable importance can and should be used to inform species distribution modeling decisions; (2) Species-environment relationships derived from small-scale studies can be applied (if ecologically justified) to regional scales for the purposes of variable selection for species distribution modelling; (3) The use of distal predictors in SDMs can be justified; however, the interpretation of such predictors may be limited by confounding variables and their suitability will likely be study specific; (4) Temporal windows and descriptive statistics (e.g. mean, cumulative sum, maximum) for continuous environmental predictors should be selected to match the ecology and life stage of the focal organism; (5) Data constraints may affect the performance of candidate predictor variables in SDMs (e.g. sample size, lack of proximal data), and may also influence decisions regarding the set of candidate predictor variables considered (e.g. characteristics of available predictor data). We recommend that future studies on species distribution modelling consider these conclusions and follow our proposed framework (Fig. 5) to improve the selection of variables for inclusion in SDMs. This framework is particularly relevant when multiple predictors (or predictor metrics) are hypothesised to affect the focal organism and sufficient data on species occurrence and predictors are available to conduct a detailed exploration of their relationships.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00442-022-05110-1>.

Acknowledgements We thank K. Arpin, C. Attridge, A. Balbar, K. Filbee-Dexter, J. Lindley, M. Neufeld, R.E. Scheibling, A. Schmidt, and C. Snow for their assistance with field work. We also thank K. Krumhansl, W. Lambert, M. McCuller, J. O'Brien, M. Saunders, and R.E. Scheibling for providing raw data and answering questions about their data for use in this study. A. Bates, E. Oliver, and two anonymous reviewers provided feedback on an earlier version of this manuscript.

Author contribution statement CJP collected and analysed field data, developed the study design, conducted data analysis, and wrote the manuscript. AM and DD collected field data, supervised the study design and analysis, and edited the manuscript.

Funding This research was funded by the Natural Sciences and Engineering Research Council (NSERC) Canadian Healthy Oceans Network and its Partners: Department of Fisheries and Oceans Canada and INREST (representing the Port of Sept-Îles and City of Sept-Îles), an NSERC Discovery Grant to AM, an OFI Seed Fund to DD and AM, and an NSERC Canada Graduate Scholarship, a Nova Scotia Graduate Scholarship to C.J.P.

Availability of data and material The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability The R code used during the current study is available from the corresponding author on reasonable request.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethics approval All applicable national and institutional guidelines for the care and use of animals were followed in accordance with the recommendations of the Canadian Council on Animal Care and the policies of Dalhousie University.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Arkema KK (2009) Flow-mediated feeding in the field: consequences for the performance and abundance of a sessile marine invertebrate. *Mar Ecol Prog Ser* 388:207–220. <https://doi.org/10.3354/meps08140>
- Asch RG, Erismann B (2018) Spawning aggregations act as a bottleneck influencing climate change impacts on a critically endangered reef fish. *Divers Distrib* 24:1712–1728. <https://doi.org/10.1111/ddi.12809>
- Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol Model* 200:1–19. <https://doi.org/10.1016/j.ecolmodel.2006.07.005>
- Austin MP, Van Niel KP (2011) Improving species distribution models for climate change studies: variable selection and scale. *J Biogeogr* 38:1–8. <https://doi.org/10.1111/j.1365-2699.2010.02416.x>
- Barbet-Massin M, Jetz W (2014) A 40-year, continent-wide, multi-species assessment of relevant climate predictors for species distribution modelling. *Divers Distrib* 20:1285–1295. <https://doi.org/10.1111/ddi.12229>
- Barton K (2019) MuMIn: Multi-Model Inference. R package version 1.43.15. <https://CRAN.R-project.org/package=MuMIn>
- Bayer MM, Todd CD, Hoyle JE, Wilson JF (1997) Wave-related abrasion induces formation of extended spines in a marine bryozoan. *Proc R Soc B* 264:1605–1611. <https://doi.org/10.1098/rspb.1997.0224>
- Berman J, Harris L, Lambert W et al (1992) Recent invasions of the Gulf of Maine: three contrasting ecological histories. *Conserv Biol* 6:435–441. <https://doi.org/10.1046/j.1523-1739.1992.06030.435.x>
- Blanco A, Larrinaga AR, Neto JM et al (2021) Spotting intruders: species distribution models for managing invasive intertidal macroalgae. *J Environ Manag* 281:111861. <https://doi.org/10.1016/j.jenvman.2020.111861>
- Brodie SJ, Thorson JT, Carroll G et al (2020) Trade-offs in covariate selection for species distribution models: a methodological comparison. *Ecography* 43:11–24. <https://doi.org/10.1111/ecog.04707>
- Brooks ME, Kristensen K, van Benthem KJ et al (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J* 9:378–400. <https://doi.org/10.3929/ethz-b-000240890>
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
- Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. *Soc Methods Res* 33:261–304. <https://doi.org/10.1177/0049124104268644>
- Caines S, Gagnon P (2012) Population dynamics of the invasive bryozoan *Membranipora membranacea* along a 450-km latitudinal range in the subarctic northwestern Atlantic. *Mar Biol* 159:1817–1832. <https://doi.org/10.1007/s00227-012-1972-z>
- Castillo KD, Lima FP (2010) Comparison of in situ and satellite-derived (MODIS-Aqua/Terra) methods for assessing temperatures on coral reefs. *Limnol Oceanogr: Methods* 8:107–117. <https://doi.org/10.4319/lom.2010.8.0107>
- Clarke A, Griffiths HJ, Barnes DK et al (2009) Spatial variation in seabed temperatures in the Southern Ocean: implications for benthic ecology and biogeography. *J Geophys Res* 114:G03003. <https://doi.org/10.1029/2008JG000886>
- Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Erlbaum, Hillsdale
- de Costa MDP, Muelbert JH, Vieira JP, Castello JP (2015) Dealing with temporal variation and different life stages of whitemouth croaker *Micropogonias furnieri* (Actinopterygii, Sciaenidae) in species distribution modeling to improve essential estuarine fish habitat identification. *Hydrobiologia* 762:195–208. <https://doi.org/10.1007/s10750-015-2348-4>
- Denley D, Metaxas A (2016) Quantifying mortality of modular organisms: a comparison of partial and whole-colony mortality in a colonial bryozoan. *Ecosphere* 7:e01483. <https://doi.org/10.1002/ecs2.1483>
- Denley D, Metaxas A (2017) Lack of substrate specificity contributes to invasion success and persistence of *Membranipora membranacea* in the northwest Atlantic. *Mar Ecol Prog Ser* 580:117–129. <https://doi.org/10.3354/meps12287>
- Denley D, Metaxas A, Fennel K (2019a) Community composition influences the population growth and ecological impact of invasive species in response to climate change. *Oecologia* 189:537–548. <https://doi.org/10.1007/s00442-018-04334-4>
- Denley D, Metaxas A, Simard N (2019b) Ocean temperature does not limit the establishment and rate of secondary spread of an ecologically significant invasive bryozoan in the northwest Atlantic. *Aquat Invasions*. <https://doi.org/10.3391/ai.2019.14.4.03>

- DeWeber JT, Wagner T (2018) Probabilistic measures of climate change vulnerability, adaptation action benefits, and related uncertainty from maximum temperature metric selection. *Glob Chang Biol* 24:2735–2748. <https://doi.org/10.1111/gcb.14101>
- DFO (2011) Aquatic invasive species: coffin box in newfoundland and labrador waters. Fisheries and Oceans Canada
- Dormann CF (2007) Promising the future? Global change projections of species distributions. *Basic Appl Ecol* 8:387–397. <https://doi.org/10.1016/j.baec.2006.11.001>
- Dormann CF, Elith J, Bacher S et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Duffy GA, Chown SL (2017) Explicitly integrating a third dimension in marine species distribution modelling. *Mar Ecol Prog Ser* 564:1–8. <https://doi.org/10.3354/meps12011>
- Eckman JE (1990) A model of passive settlement by planktonic larvae onto bottoms of differing roughness. *Limnol Oceanogr* 35:887–901. <https://doi.org/10.4319/lo.1990.35.4.0887>
- Eckman JE, Duggins DO, Sewell AT (1989) Ecology of under story kelp environments. I. Effects of kelps on flow and particle transport near the bottom. *J Exp Mar Biol Ecol* 129:173–187. [https://doi.org/10.1016/0022-0981\(89\)90055-5](https://doi.org/10.1016/0022-0981(89)90055-5)
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Semant* 40:677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Filbee-Dexter K, Feehan C, Scheibling R (2016) Large-scale degradation of a kelp ecosystem in an ocean warming hotspot. *Mar Ecol Prog Ser* 543:141–152. <https://doi.org/10.3354/meps11554>
- Fourcade Y, Besnard AG, Secondi J (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob Ecol Biogeogr* 27:245–256. <https://doi.org/10.1111/geb.12684>
- Franklin J (2013) Species distribution models in conservation biogeography: developments and challenges. *Divers Distrib* 19:1217–1223. <https://doi.org/10.1111/ddi.12125>
- Gaines S, Brown S, Roughgarden J (1985) Spatial variation in larval concentrations as a cause of spatial variation in settlement for the barnacle, *Balanus glandula*. *Oecologia* 67:267–272. <https://doi.org/10.1007/BF00384297>
- Galipaud M, Gillingham MA, Dechaume-Moncharmont F-X (2017) A farewell to the sum of Akaike weights: the benefits of alternative metrics for variable importance estimations in model selection. *Methods Ecol Evol* 8:1668–1678. <https://doi.org/10.1111/2041-210X.12835>
- Gallardo B, Zieritz A, Aldridge DC (2015) The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine invaders. *PLoS ONE* 10:e0125801. <https://doi.org/10.1371/journal.pone.0125801>
- Gardner AS, Maclean IM, Gaston KJ (2019) Climatic predictors of species distributions neglect biophysiological meaningful variables. *Divers Distrib* 25:1318–1333. <https://doi.org/10.1111/ddi.12939>
- Gelman A (2008) Scaling regression inputs by dividing by two standard deviations. *Stat Med* 27:2865–2873. <https://doi.org/10.1002/sim.3107>
- Graham MH (2003) Confronting multicollinearity in ecological multiple regression. *Ecology* 84:2809–2815. <https://doi.org/10.1890/02-3114>
- Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: challenges and solutions. *J Evol Biol* 24:699–711. <https://doi.org/10.1111/j.1420-9101.2010.02210.x>
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8:993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models: with applications in R. Cambridge University Press
- Hartig F (2021) DHARMa: Residual diagnostics for hierarchical (Multi-Level/Mixed) regression models. R package version 0.4.4. <https://CRAN.R-project.org/package=DHARMa>
- Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35:276–288. <https://doi.org/10.1111/j.1600-0587.2011.06999.x>
- Hiddink JG, Burrows MT, García Molinos J (2015) Temperature tracking by North Sea benthic invertebrates in response to climate change. *Glob Chang Biol* 21:117–129. <https://doi.org/10.1111/gcb.12726>
- Jarnevich CS, Stohlgren TJ, Kumar S et al (2015) Caveats for correlative species distribution modeling. *Ecol Inform* 29:6–15. <https://doi.org/10.1016/j.ecoinf.2015.06.007>
- Koehl MRA (2007) Mini review: hydrodynamics of larval settlement into fouling communities. *Biofouling* 23:357–368. <https://doi.org/10.1080/08927010701492250>
- Krumhansl K, Scheibling R (2011) Detrital production in Nova Scotian kelp beds: patterns and processes. *Mar Ecol Prog Ser* 421:67–82. <https://doi.org/10.3354/meps08905>
- Krumhansl KA, Lee JM, Scheibling RE (2011) Grazing damage and encrustation by an invasive bryozoan reduce the ability of kelps to withstand breakage by waves. *J Exp Mar Biol Ecol* 407:12–18. <https://doi.org/10.1016/j.jembe.2011.06.033>
- La Marca W, Elith J, Firth RS et al (2019) The influence of data source and species distribution modelling method on spatial conservation priorities. *Divers Distrib* 25:1060–1073. <https://doi.org/10.1111/ddi.12924>
- Lambert W, Levin P, Berman J (1992) Changes in the structure of a New England (USA) kelp bed: the effects of an introduced species? *Mar Ecol Prog Ser* 88:303–307. <https://doi.org/10.3354/meps088303>
- Leroux SJ (2019) On the prevalence of uninformative parameters in statistical models applying model selection in applied ecology. *PLoS ONE* 14:e0206711. <https://doi.org/10.1371/journal.pone.0206711>
- Metaxas A, Saunders M (2009) Quantifying the “bio-” components in biophysical models of larval transport in marine benthic invertebrates: advances and pitfalls. *Biol Bull* 216:257–272. <https://doi.org/10.1086/BBLv216n3p257>
- Mod HK, Scherrer D, Luoto M, Guisan A (2016) What we use is not what we know: environmental predictors in plant distribution models. *J Veg Sci* 27:1308–1322. <https://doi.org/10.1111/jvs.12444>
- Møller A, Jennions MD (2002) How much variance can be explained by ecologists and evolutionary biologists? *Oecologia* 132:492–500. <https://doi.org/10.1007/s00442-002-0952-2>
- Nakagawa S, Johnson PC, Schielzeth H (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface* 14:20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Neuheimer AB, Taggart CT (2007) The growing degree-day and fish size-at-age: the overlooked metric. *Can J Fish Aquat Sci* 64:375–385. <https://doi.org/10.1139/f07-003>
- NOAA Coral Reef Watch (2020) NOAA coral reef watch daily global satellite sea surface temperature (a.k.a. ‘CoralTemp’) Version 3.1: Jan. 1, 1986 - Oct. 21, 2019
- Ospina R, Ferrari SL (2012) A general class of zero-or-one inflated beta regression models. *Comput Stat Data Anal* 56:1609–1623. <https://doi.org/10.1016/j.csda.2011.10.005>
- Pandori LL, Sorte CJ (2019) The weakest link: sensitivity to climate extremes across life stages of marine invertebrates. *Oikos* 128:621–629. <https://doi.org/10.1111/oik.05886>

- Pawlik JR, Butman CA (1993) Settlement of a marine tube worm as a function of current velocity: interacting effects of hydrodynamics and behavior. *Limnol Oceanogr* 38:1730–1740. <https://doi.org/10.4319/lo.1993.38.8.1730>
- Pearson RG (2007) Species' distribution modeling for conservation educators and practitioners. American Museum of Natural History, New York
- Peterson AT, Soberón J, Pearson RG et al (2011) Ecological niches and geographic distributions. Princeton University Press, Princeton, New Jersey
- Petitpierre B, Broennimann O, Kueffer C et al (2017) Selecting predictors to maximize the transferability of species distribution models: Lessons from cross-continental plant invasions. *Glob Ecol Biogeogr* 26:275–287. <https://doi.org/10.1111/geb.12530>
- Pineda MC, McQuaid CD, Turon X et al (2012) Tough adults, frail babies: an analysis of stress sensitivity across early life-history stages of widely introduced marine invertebrates. *PLoS ONE* 7:e46672. <https://doi.org/10.1371/journal.pone.0046672>
- Porfiro LL, Harris RM, Lefroy EC et al (2014) Improving the use of species distribution models in conservation planning and management under climate change. *PLoS ONE* 9:e113749. <https://doi.org/10.1371/journal.pone.0113749>
- Pörtner HO, Peck L, Somero G (2007) Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view. *Philos Trans R Soc B* 362:2233–2258. <https://doi.org/10.1098/rstb.2006.1947>
- Pratt MC (2008) Living where the flow is right: how flow affects feeding in bryozoans. *Integr Comp Biol* 48:808–822. <https://doi.org/10.1093/icb/icn052>
- Putnam HM, Edmunds PJ, Fan T-Y (2010) Effect of a fluctuating thermal regime on adult and larval reef corals. *Invertebr Biol* 129:199–209. <https://doi.org/10.1111/j.1744-7410.2010.00199.x>
- R Core Team (2020) R: A language and environment for statistical computing. Version 4.0.3. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Reglero P, Tittensor DP, Álvarez-Berastegui D et al (2014) World-wide distributions of tuna larvae: revisiting hypotheses on environmental requirements for spawning habitats. *Mar Ecol Prog Ser* 501:207–224. <https://doi.org/10.3354/meps10666>
- Richards SA (2005) Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology* 86:2805–2814. <https://doi.org/10.1890/05-0074>
- Richards SA (2008) Dealing with overdispersed count data in applied ecology: Overdispersed count data. *J Appl Ecol* 45:218–227. <https://doi.org/10.1111/j.1365-2664.2007.01377.x>
- Richards SA, Whittingham MJ, Stephens PA (2011) Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. *Behav Ecol Sociobiol* 65:77–89. <https://doi.org/10.1007/s00265-010-1035-8>
- Robinson LM, Elith J, Hobday AJ et al (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob Ecol Biogeogr* 20:789–802. <https://doi.org/10.1111/j.1466-8238.2010.00636.x>
- RStudio Team (2020) RStudio: integrated development environment for R. Version 1.3.1093. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- Santika T, Hutchinson MF (2009) The effect of species response form on species distribution model prediction and inference. *Ecol Model* 220:2365–2379. <https://doi.org/10.1016/j.ecolmodel.2009.06.004>
- Saunders M, Metaxas A (2007) Temperature explains settlement patterns of the introduced bryozoan *Membranipora membranacea* in Nova Scotia, Canada. *Mar Ecol Prog Ser* 344:95–106. <https://doi.org/10.3354/meps06924>
- Saunders M, Metaxas A (2008) High recruitment of the introduced bryozoan *Membranipora membranacea* is associated with kelp bed defoliation in Nova Scotia, Canada. *Mar Ecol Prog Ser* 369:139–151. <https://doi.org/10.3354/meps07669>
- Saunders MI, Metaxas A (2009a) Population dynamics of a nonindigenous epiphytic bryozoan *Membranipora membranacea* in the western North Atlantic: effects of kelp substrate. *Aquat Biol* 8:83–94. <https://doi.org/10.3354/ab00208>
- Saunders MI, Metaxas A (2009b) Effects of temperature, size, and food on the growth of *Membranipora membranacea* in laboratory and field studies. *Mar Biol* 156:2267–2276. <https://doi.org/10.1007/s00227-009-1254-6>
- Saunders MI, Metaxas A, Filgueira R (2010) Implications of warming temperatures for population outbreaks of a nonindigenous species (*Membranipora membranacea*, Bryozoa) in rocky subtidal ecosystems. *Limnol Oceanogr* 55:1627–1642. <https://doi.org/10.4319/lo.2010.55.4.1627>
- Scheibling R, Gagnon P (2009) Temperature-mediated outbreak dynamics of the invasive bryozoan *Membranipora membranacea* in Nova Scotian kelp beds. *Mar Ecol Prog Ser* 390:1–13. <https://doi.org/10.3354/meps08207>
- Scheibling RE, Hennigar AW, Balch T (1999) Destructive grazing, epiphytism, and disease: the dynamics of sea urchin–kelp interactions in Nova Scotia. *Can J Fish Aquat Sci* 56:15. <https://doi.org/10.1139/f99-163>
- Schiel DR, Steinbeck JR, Foster MS (2004) Ten years of induced ocean warming causes comprehensive changes in marine benthic communities. *Ecology* 85:1833–1839. <https://doi.org/10.1890/03-3107>
- Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. *Methods Ecol Evol* 1:103–113. <https://doi.org/10.1111/j.2041-210X.2010.00012.x>
- Schwaninger HR (2008) Global mitochondrial DNA phylogeography and biogeographic history of the antitropically and longitudinally disjunct marine bryozoan *Membranipora membranacea* L. (Cheilostomata): another cryptic marine sibling species complex? *Mol Phylogenet Evol* 49:893–908. <https://doi.org/10.1016/j.ympev.2008.08.016>
- Stobart B, Mayfield S, Mundy C et al (2016) Comparison of in situ and satellite sea surface-temperature data from South Australia and Tasmania: how reliable are satellite data as a proxy for coastal temperatures in temperate southern Australia? *Mar Freshw Res* 67:612–625. <https://doi.org/10.1071/MF14340>
- Stricker SA, Reed CG, Zimmer RL (1988) The cyphonautes larva of the marine bryozoan *Membranipora membranacea*. I. General morphology, body wall, and gut. *Can J Zool* 66:368–383. <https://doi.org/10.1139/z88-054>
- Symonds MR, Moussalli A (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav Ecol Sociobiol* 65:13–21. <https://doi.org/10.1007/s00265-010-1037-6>
- Synes NW, Osborne PE (2011) Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Glob Ecol Biogeogr* 20:904–914. <https://doi.org/10.1111/j.1466-8238.2010.00635.x>
- Wagenmakers E-J, Farrell S (2004) AIC model selection using Akaike weights. *Psychon Bull Rev* 11:192–196. <https://doi.org/10.3758/BF03206482>
- Watanabe S, Scheibling RE, Metaxas A (2010) Contrasting patterns of spread in interacting invasive species: *Membranipora membranacea* and *Codium fragile* off Nova Scotia. *Biol Invas* 12:2329–2342. <https://doi.org/10.1007/s10530-009-9647-5>
- Wernberg T, Krumhansl K, Filbee-Dexter K, Pedersen MF (2019) Status and trends for the world's kelp forests. In: Sheppard C (ed) World seas: an environmental evaluation, 2nd edn. Elsevier, London, pp 57–78
- Wilbur HM (1980) Complex life cycles. *Annu Rev Ecol Syst* 11:67–93. <https://doi.org/10.1146/annurev.es.11.110180.000435>

- Williams KJ, Belbin L, Austin MP et al (2012) Which environmental variables should I use in my biodiversity model? *Int J Geogr Inform Sci* 26:2009–2047. <https://doi.org/10.1080/13658816.2012.698015>
- Yorke A, Metaxas A (2012) Relative importance of kelps and fucoids as substrata of the invasive epiphytic bryozoan *Membranipora membranacea* in Nova Scotia, Canada. *Aquat Biol* 16:17–30. <https://doi.org/10.3354/ab00419>
- Yoshioka PM (1982) Role of planktonic and benthic factors in the population dynamics of the bryozoan *Membranipora membranacea*. *Ecology* 63:457–468. <https://doi.org/10.2307/1938963>