




## Original Article

# Assessing effects of genetic, environmental, and biotic gradients in species distribution modelling

J. Benjamin Lowen <sup>1\*</sup>, Devorah R. Hart<sup>2</sup>, Ryan R. E. Stanley<sup>1</sup>, Sarah J. Lehnert<sup>1</sup>, Ian R. Bradbury<sup>3</sup>, and Claudio DiBacco<sup>1</sup>

<sup>1</sup>Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia B2Y 4A2, Canada

<sup>2</sup>NMFS Northeast Fisheries Science Center, 166 Water St., Woods Hole, MA, USA

<sup>3</sup>Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John's, Newfoundland A1C 5X1, Canada

\*Corresponding author: tel: +1 902 426 1865; e-mail: [ben.lowen@dfo-mpo.gc.ca](mailto:ben.lowen@dfo-mpo.gc.ca).

Lowen, J. B., Hart, D. R., Stanley, R. R. E., Lehnert, S. J., Bradbury, I. R., and DiBacco, C. Assessing effects of genetic, environmental, and biotic gradients in species distribution modelling. – ICES Journal of Marine Science, 76: 1762–1775.

Received 14 November 2018; revised 19 February 2019; accepted 3 March 2019; advance access publication 8 April 2019.

To develop more reliable marine species distribution models (SDMs), we examine how genetic, climatic, and biotic interaction gradients give rise to prediction error in marine SDM. Genetic lineages with distinct ecological requirements spanning genetic gradients have yet to be treated separately in marine SDM, which are often constrained to modeling the potential distribution of one biological unit (e.g. lineage or species) at a time. By comparing SDM performance for the whole species or where observation and predictions were partitioned among geographically discontinuous genetic lineages, we first identified the appropriate biological unit for modeling sea scallop. Prediction errors, in particular contiguous omissions at the northern range margins were effectively halved in genetic lineage SDM (Total error=15%) versus whole species SDM. Remaining SDM prediction error was strongly associated with: i) Sharp climatic gradients (abrupt and persistent spatial shifts in limiting temperatures) found within continental shelf breaks and bottom channels. ii) A biotic gradient in the predation of sea scallop juveniles by the sand star within the Hudson Shelf USA. Our findings highlight how the accuracy of marine SDM is dependent on capturing the appropriate biological unit for modeling (e.g. lineages rather than species) and adequately resolving limiting abiotic and biotic interaction gradients.

**Keywords:** biotic gradients, environmental gradients, genetic structure, marine species distribution, prediction error, seascape, species distribution modelling

## Introduction

Accurately predicting and understanding changes in marine species distributions using species distribution models (SDMs) has become an increasingly important topic in ecology, evolutionary biology, and biogeography (May and MacArthur, 1972; Chase and Leibold, 2003). Reliable predictions of suitable habitat in marine environments are required for identifying present-day and future distributions of ecologically, economically, and culturally important species. Indeed, accurate predictions of habitat have proven invaluable for the successful conservation of sensitive or at risk species and the management of aquatic invasive species (AIS) (Therriault and Herborg, 2008).

To develop more reliable marine SDMs, we examine how genetic, climatic, and biotic interaction gradients can give rise to prediction error in marine SDM. Recent increases in the application of SDM in the marine environment have coincided with increased availability of high-resolution species occurrence and environmental data (Tyberghein *et al.*, 2012; Lowen *et al.*, 2016; Lowen and DiBacco, 2017), in conjunction with the development of new software platforms and algorithms to analyse and synthesize these data (Franklin, 2010; Phillips *et al.*, 2017), and statistical methods to optimize SDM predictive utility by reducing model overfitting and spatial bias (Muscarella *et al.*, 2014; Phillips *et al.*, 2017). Despite this progress, further advances in marine SDM utility may yet depend on models that simultaneously capture the

appropriate biological unit for modelling and adequately resolve limiting abiotic and biotic gradients that drive species distributions (Franklin, 2010; Lowen *et al.*, 2016). Here, the biological unit can vary from species to genetically distinct sub-populations or lineages that effectively stem from genetic gradients or clines, yet marine SDM are currently based solely on the species as the biological unit of interest despite evidence to the contrary from terrestrial environments (Hernandez *et al.*, 2006).

Genetic gradients can co-vary with the environment to form spatially and ecologically disjunct lineages that may represent the biological units that matter most to understand the mechanisms that shape species distributions in terrestrial environments (Hernandez *et al.*, 2006). To date, SDM in marine environments have only focused on the species as the biological unit of interest, but this approach may lead to spurious predictions when a species is composed of ecologically distinct sub-populations or lineages that exhibit local specialization and occupy a distinct fundamental niche (Valladares *et al.*, 2014). Individual SDM effectively predict the potential distribution of one biological unit (e.g. lineage or species) at a time (Franklin, 2010) and cannot distinguish multiple biological units with distinct ecological requirements simultaneously (Davis and Shaw, 2001; Fitzpatrick and Keller, 2015). In addition, multispecies SDM that can model the ecological response and potential distribution of more than one biological unit simultaneously have yet to be applied to model ecologically distinct lineages separately (Franklin, 2010; Pollock *et al.*, 2014). In terrestrial environments, prediction error associated with a decoupling of the potential distribution in whole species vs. lineage level SDM generally occurs at a species' environmental range margins where ecologically specialized lineages are more tolerant of environmental extremes (Romero *et al.*, 2014; Valladares *et al.*, 2014).

Species in terrestrial environments are in effect highly heterogeneous entities as a result of environmentally driven demographic and/or adaptive processes (Hernandez *et al.*, 2006). Consequently there is a growing consensus that terrestrial species should be modelled as subunits based on distinct genetic lineages or as recognized subspecies (Hernandez *et al.*, 2006). Identifying such heterogeneity in the marine realm is often challenged by an absence of conspicuous barriers to gene flow, which often leads to inappropriate assumptions of panmixia. However, with the advent of new genomic approaches to evaluate population structure, it has become increasingly evident that intraspecific variation can lead to variability in habitat requirements across species ranges in the marine environment (Stanley *et al.*, 2018). Indeed, climatic breaks can represent major, albeit cryptic, barriers to gene flow, and population connectivity that in turn shape spatial patterns in genetic structure along genetic gradients (Galarza *et al.*, 2009; Jeffery *et al.*, 2018; Lehnert *et al.*, 2018). It follows that where species span genetic–environmental gradients in the marine environment that effectively split species into ecologically distinct populations or sub-units, the biological unit considered for modelling should reflect genetic structure (Hernandez *et al.*, 2006).

Numerous studies have identified limiting abiotic or biotic–environmental gradients as important drivers of marine SDM prediction error in areas where these gradients are not well resolved (Franklin, 2010; Lowen *et al.*, 2016). One advantage of identifying error attributable to genetic structure as in this study is that prediction error associated with not resolving limiting abiotic or biotic gradients can be more accurately pinpointed and

partitioned. For example, climatological temperature and salinity gradients within oceanic fronts that limit marine species distributions over relatively fine spatial scales (metres to kilometres) (Cromwell and Reid, 1956; Sbrocco and Barber, 2013) must be sufficiently resolved to allow accurate prediction in SDM (Davies and Guinotte, 2011; Basher *et al.*, 2014; Lowen *et al.*, 2016). Found in areas of coastal upwelling, areas of high topographic complexity (shelf breaks, channels, or in the near shore land–sea interface), and near freshwater discharge plumes, sharp climatic gradients associated with oceanic fronts are ubiquitous in the marine environment (Tyberghein *et al.*, 2012; Sbrocco and Barber, 2013). With the recent availability of high-resolution climatological data layers for use in SDM, considerable progress has been made towards providing readily available environmental data that can characterize and capture fine-scale variation in temperature and salinity at the surface of these oceanic fronts (Lowen *et al.*, 2016). Nonetheless, climatological temperature and salinity data layers that cover the range of broadly distributed (>800 km) benthic species occupying deeper waters (<10 m) are currently too coarse (>5 km resolution) to adequately capture very fine-scale limiting variations in temperature and salinity (100' sm–5 km) within the deeper waters associated with these oceanic fronts (Tyberghein *et al.*, 2012; Lowen *et al.*, 2016).

Likewise, missing biological covariates linked to biotic interaction gradients in environmental datasets in SDM are also a potentially important source of SDM prediction error in marine environments (Davies and Guinotte, 2011; Basher *et al.*, 2014; Lowen *et al.*, 2016). The presence of heterospecifics can lead to gradients in predation pressure and/or competition for resources that also shape species distributions (Franklin, 2010). Not resolving predation or competitive interactions in SDM may result in overprediction of a species distribution in areas where predation or competition negatively influence species distribution.

The Atlantic sea scallop (*Placopecten magellanicus*) represents a useful case study species with which to account for SDM prediction error arising from genetic, climatic, and biotic interaction gradients in a heterogeneous marine environment. The distribution of this economically important species has been well documented for decades via commercial stock surveys throughout its northwest Atlantic range spanning latitudes of ~35 to 65°N (Hart and Chute, 2004; NEFSC, 2014; DFO, 2016). These empirically based, high-resolution distributional data represent a unique opportunity to validate and partition sources of error in SDM (Franklin, 2010). The scallop range spans a vast latitudinal area from North Carolina, USA, to Newfoundland (NL), Canada where the cold Labrador Current meets the warm Gulf Stream. This convergence leads to large gradients in ocean temperature and salinity, that drive genetic differentiation along a genetic gradient (cline) where up to three ecologically distinct lineages have been identified (Van Wyngaarden *et al.*, 2017; Lehnert *et al.*, 2018). This genetic gradient presents an opportunity to test the hypothesis that marine species distributions can be better predicted with lower SDM prediction error, when genetic structure is incorporated. Sharp climatic gradients (abrupt shifts in temperature) that limit sea scallop establishment associated with continental shelf breaks and bottom topographic channels (Cromwell and Reid, 1956; Sbrocco and Barber, 2013) could also drive prediction errors in sea scallop SDM where these climatic gradients are poorly resolved. Sea scallop may also compete with mussels and tunicates for food space, while spat and juveniles are very vulnerable to predation by certain crabs (cancer spp.), lobster

(*Homarus americanus*), gastropods, and sea stars (*Asterias* and *Astropecten* spp.) and ground fish (Hart and Chute, 2004). Predation of sea scallop juveniles by the sand star *Astropecten americanus* in deeper and colder waters around shelf breaks and bottom channels in the Mid-Atlantic Bight (MAB) has been identified as an important factor affecting the fine-scale distribution in sea scallop (Hart, 2006; Shank et al., 2012) not captured by SDM using only climatic drivers of distribution.

To account for patterns in prediction error in marine SDM stemming from genetic, climatic, and biotic gradients, we first modelled sea scallop distribution at the species level (all known occurrences) and documented clusters of prediction error as a baseline. After which, using known population structure that delineates up to three genetic lineages (Lehnert et al., 2018), we tested whether multiple SDMs run independently for each lineage more accurately predicted scallop distribution than range-wide (RW) species SDM. Accuracy of SDM was assessed against the known distribution of sea scallop, including commercial stocks characterized annually over the past 20 years (Hart and Chute, 2004; NEFSC, 2014; DFO, 2016). After identifying the optimal biological unit (i.e. lineage or species?) with which to predict sea scallop distribution in SDM we could subsequently evaluate the role of climatological and biotic gradients in driving patterns of remaining prediction error. Specifically, we identified; (i) whether or not currently available resolutions of abiotic–environmental data are too coarse to capture sharp environmental gradients around shelf breaks and topographic features (i.e. channels) that might limit sea scallop establishment across its known range, and (ii) whether the distribution and abundance of the predatory sand star *A. americanus* (Hart, 2006; Shank et al., 2012) also influences SDM accuracy. SDMs are increasingly being used to predict and understand changes in marine species distributions and this study highlights how the accuracy of marine SDM can be dependent on capturing the appropriate biological unit for modelling (e.g. genetic lineages rather than species) and adequately resolving limiting abiotic and biotic gradients that can drive clusters of prediction error in areas where they are poorly resolved.

## Methods

### Accounting for SDM prediction error across genetic gradients?

#### *Distribution of occurrence data, commercial stocks, and genetic lineages*

Two sea scallop distributional datasets were employed. The first was used for MaxEnt SDM development and based only on collated presence locations which were spatially rarefied to prevent spatial autocorrelation as below (“Model fitting procedure and evaluation”). Sea scallop presence locations encompassing non-commercial and commercial populations or stocks across its known distribution from 1998 to 2012 were collated from datasets collected by Fisheries and Oceans Canada (DFO, 2016) and the National Oceanic and Atmospheric Administration (Hart and Chute, 2004; NEFSC, 2014).

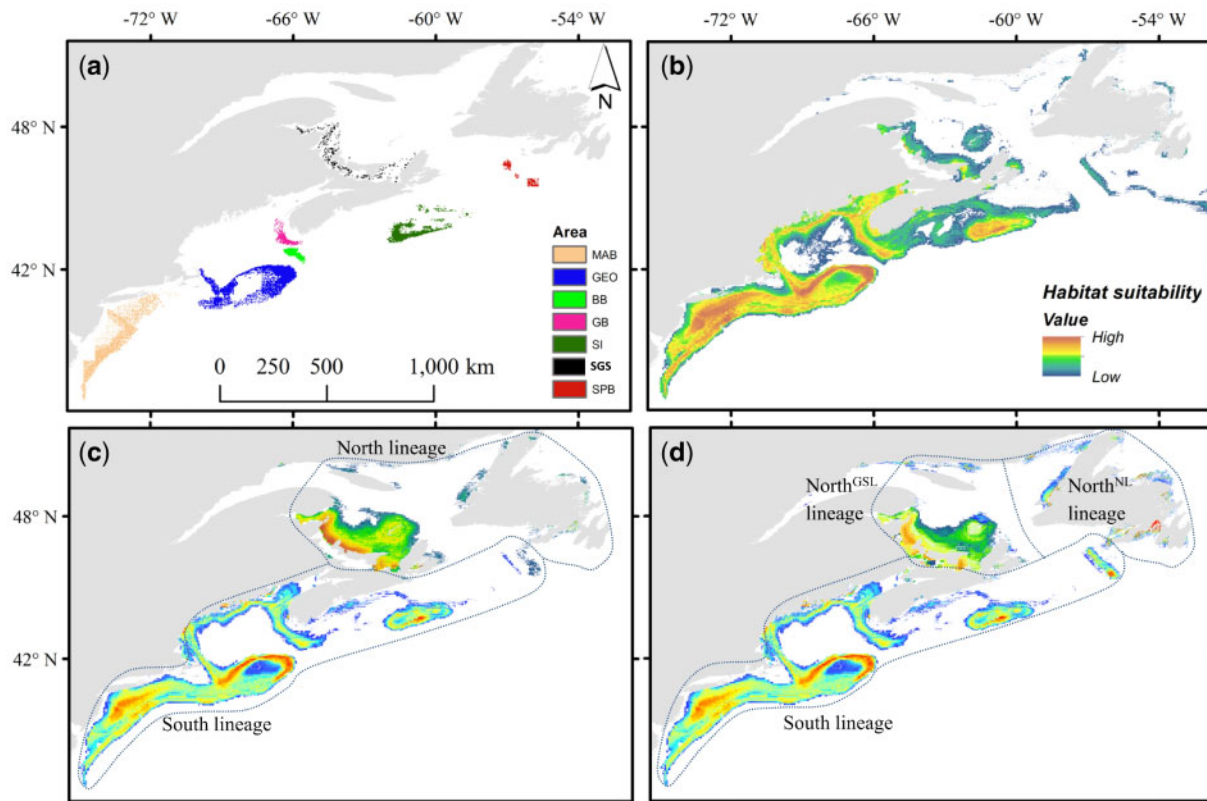
The second distributional dataset was subsequently used to validate SDM presence–absence predictions against known presence–absences within commercial scallop stocks; thereby providing a quasi-independent external assessment of SDM performance as in “Model fitting procedure and evaluation”. Presence and absence data, collated from stock assessments spanning 1998 to 2016 (NEFSC, 2014; DFO, 2016), was obtained for

seven commercial stocks which have been well characterized over the past 20 years due to their commercial importance (Figure 1a). These commercial stocks included the: MAB, Georges Bank (GEO), Browns Bank (BB), German Bank (GB), Sable Island Bank (SI), southern Gulf stock (SGS), and St. Pierre Bank (SPB) stocks (Figure 1a). The design of the stock assessment trawl surveys is outlined in detail in Davies et al. (2014). These surveys are based on a random stratified design in which the ocean is divided into zones, or strata, of similar depth and habitat, with dredge and camera samples taken randomly within these zones. Samples were collected using a modified commercial 8-foot sea scallop dredge. Survey data are subsequently standardized to a tow area of 1951 m<sup>2</sup> tow<sup>-1</sup> in Canadian and 4500 m<sup>2</sup> in the US waters.

The sea scallop genetic structure incorporated in this study is based on a northern vs. southern genetic lineage split distinguished by a primary cline described by Van Wyngaarden et al. (2017) and Lehnert et al. (2018), hereafter referred to as “north” and “south” SDM lineages. The north lineage was subsequently split along a secondary cline, which distinguished scallop within the Gulf of St. Lawrence (GSL) from coastal NL. Northern lineages based on the secondary cline are hereafter referred to as “North<sup>GSL</sup>” and “North<sup>NL</sup>” SDM lineages. To assess the relative importance of treating these lineages separately, we employed three SDM scenarios based on the environmental predictors outlined below as in “Model fitting procedure and evaluation”. Scenario 1 treated all known occurrences of scallop as a single population in a RW species SDM ( $n = 2779$  occurrences) (Figure 1b). Scenario 2 modelled northern ( $n = 354$  occurrences) and southern ( $n = 2420$  occurrences) lineages corresponding with the primary cline separately (Figure 1c). Scenario 3 subdivided and modelled the northern lineage as two separate populations, namely North<sup>GSL</sup> ( $n = 324$  occurrences) and North<sup>NL</sup> ( $n = 30$  occurrences) lineages corresponding with the secondary cline (Figure 1d). Occurrence locations for SDM training and testing were spatially partitioned, as appropriate, to match the geographic extent of each lineage in these three scenarios. Fine-scale genetic structure was determined from sea scallop sampled across 27 sample locations spanning the sea scallops full distributional extent and all commercially important populations from ~37 to 51°N (Lehnert et al., 2018). Such extensive sampling provided a relatively fine-scale spatial approximation of genetic structure to use as a basis for spatially partitioning scallop occurrences among identified genetic clusters for SDM (Lehnert et al., 2018).

#### *Environmental predictors*

Environmental variables that are ecologically relevant and available to model the establishment of *P. magellanicus* across its known extent were employed here; specifically seasonal temperature and salinity. Benthic temperature and benthic salinity climatological data layers (1993–2012) were assembled at spatial resolutions interpretable to 8 km<sup>2</sup> from a numerical ocean general circulation model (GLORYSV2) adapted and validated against the study area by Fisheries and Oceans Canada (Z. Wang, pers. comm.) as described in Beazley et al. (2016). This reanalysis product is a model run with data assimilation and is designed to reproduce the ocean state. Unlike other numerical data that had been validated and were available for use [e.g. NEMO (Nucleus for European Modelling of the Ocean) version 2.3], GLORYSV2 benthic extent spanned both the southern and northern limits of the seas scallops distribution. Temperature and salinity data were



**Figure 1.** (a) Commercial sea scallop (*P. magellanicus*) distribution, including MAB, GEO, BB, GB, SI, SGS, and SPB stocks. Predicted distributions are based on three SDM scenarios predicted from 1993 to 2012 seasonal temperature and salinity climatologies, (b) No cline (RW species), (c) Primary cline (south vs. north lineages), and (d) Secondary cline (North<sup>GSL</sup> vs. North<sup>NL</sup> lineages). Dotted lines delineate; (c) southern and northern lineages and (d) southern, North<sup>GSL</sup> and North<sup>NL</sup> lineages. The habitat suitability scale presented in panel b applies to panels c and d.

sub-divided by season corresponding with winter (January–March), spring (April–June), summer (July–September), and fall (October–December).

#### Model fitting procedure and evaluation

Suitable habitat for *P. magellanicus* in each modelling scenario was predicted using Maximum Entropy Modeling (MaxEnt, version 3.4.0) (Phillips *et al.*, 2004). Occurrence data were contrasted with a sample of background locations and correlated with environmental predictors in MaxEnt to delineate the multidimensional range of environmental conditions that determine sea scallop at the species vs. lineage level. This machine learning algorithm is particularly efficient at handling complex interactions between response and predictor variables and can accommodate a diverse suite of distributional responses to environmental gradients (Franklin, 2010). These SDM predict species distributional persistence within the seascape from the climatological time series. For example, with un-usually favourable temperatures, a population may spread but not persist in a given location. The model domain ranged from  $-76.8$  to  $-50.7^{\circ}\text{W}$  and  $32.6$  to  $55.2^{\circ}\text{N}$ . The maximum iterations of model runs were set to 1000 to allow for model convergence, while the number of background points was set to 100 000 (Phillips *et al.*, 2004). Spatial autocorrelation, likely arising from clustering of samples around favourable habitat as determined by Moran's I-distance plot, was evident in

the original occurrence data. Occurrence records used to train and test the model were spatially rarefied as in Brown (2014) to mitigate effects of spatial autocorrelation, including inflated estimates of SDM performance through overfitting (Phillips *et al.*, 2004; Mateo *et al.*, 2010). Overfitting also arises if a model is overly complex, resulting in under-estimation of suitable conditions in projected locations. To prevent model overfitting we defined optimal model settings by iteratively replicated cross validation (package SDMEval in Muscarella *et al.*, 2014) to balance goodness-of-fit with model complexity. During internal model cross validation, spatially rarefied presence data were spatially partitioned into training and testing datasets for each of ten-fold or replicates. Test data in each fold were used to calculate the accuracy of predicted distributions after estimation of threshold-dependent metrics based on the average of all replicates (Muscarella *et al.*, 2014).

Threshold-dependent metrics included the proportion of accurately predicted presence (hereafter sensitivity) and absence data (specificity). Rates of overprediction (false-positive rate) and underprediction (false-negative rate) error, which are effectively the inverse of sensitivity and specificity respectively, are also reported alongside total error (sum of underprediction and overprediction error) (Merow *et al.*, 2013; Basher *et al.*, 2014). Underprediction error occurs when presences are incorrectly predicted as absences while overprediction error occurs when

absences are incorrectly predicted as presences. Based on model sensitivity and specificity we estimated the true skill statistic (TSS = sensitivity + specificity - 1). TSS varies from -1 to +1 where negative values and those close to zero indicate distributions that are no better than random, while values approaching +1 represent increasing agreement between observed and modelled distributions. Models considered highly accurate exhibit a TSS greater than ~0.8 (Silva *et al.*, 2014), sensitivity and specificity values >0.9, and over or underprediction errors <0.1 (Franklin, 2010). These metrics are preferable to threshold-independent metrics, including area under the curve, since accuracy is not inflated by prevalence of occurrence points, size of study region, and assumptions of equal costs of omission and commission errors (Allouche *et al.*, 2006; Lobo *et al.*, 2008; Merow *et al.*, 2013). These biases would negate useful comparisons among model scenarios tested here.

Threshold-dependent metrics described were subsequently used to validate SDM presence-absence predictions against commercial stocks (presence-absences) that fall within the geographic extent of each of the tested lineage scenarios (Figure 1a). The southern lineage SDM (south of the primary cline) included MAB + GEO + BB + GB + SI + SPB as shown in Figure 1a, while the northern lineages (North or North<sup>GSL</sup>) included SGS as shown in Figure 1a. For direct comparison to lineage SDMs, validation results based on stocks that fall north or south of the primary cline within the RW SDM are reported. The objective of this accuracy assessment was to provide a quasi-independent external validation of SDM performance to determine which scenario best predicts empirical distributions of *P. magellanicus*. Stock assessment data (observed presence and absences) were scaled to the same resolution as gridded model predictions (8 km<sup>2</sup>) for this comparison by removing duplicates within a given grid.

Environmental tolerances, in the context of how seasonal salinity and temperature climatologies shape each sea scallop distribution within each SDM scenario, were determined from the relative contribution of predictors to model performance, and predicted temperature and salinity thresholds for occurrence. MaxEnt provides the percent contribution of each variable to the final model, thereby assessing the relative contribution of each variable in each tested SDM scenario (Phillips *et al.*, 2004). Predicted thresholds for occurrence for each predictor were estimated from response curves generated by MaxEnt for each tested modelling scenario.

### Accounting for SDM prediction error across environmental gradients

Remaining SDM prediction error not accounted for by incorporation of genetic structure likely stems from abiotic or biotic interaction gradients. Consequently, we considered three main environmental drivers of error in the SDM for two representative commercial scallop stocks (MAB and GEO) (Figure 1a). The first included seasonal benthic temperature and salinity climatological data used to develop the SDMs. The second included fine-scale topographic complexity, represented by bottom slope, and curvature. Finally, the third variable included sea star density (*A. americanus*) as a proxy of predation pressure on sea scallop recruits in the MAB stock (Hart, 2006; Shank *et al.*, 2012).

Seasonal bottom temperature (BT) and salinity data were available for all scallop stocks. Topographic complexity

represented by bottom slope and curvature was derived from high-resolution (500 m<sup>2</sup>) bathymetry data (Smith and Sandwell, 1997). Bathymetric slope was measured in degrees ranging from 0° (flat) to 90° (near vertical). Terrain curvature represents whether a raster cell is on a hill (negative values) or in a valley (positive values). Terrain curvature was calculated as the “slope of the slope” for a given cell based on its eight surrounding neighbours using the ARCMAP curvature tool. Data on the abundance of the sand star *A. americanus* were collected on scallop dredge surveys in the MAB (Hart, 2006; Shank *et al.*, 2012). These data were incorporated as a plausible biotic factor influencing SDM overprediction error. The impact of sand stars on sea scallop recruitment is likely greatest in areas of relatively colder water temperatures (particularly winter minima) within the MAB that represent more favourable habitat for sand stars (Hart, 2006). The prevailing hypothesis is that sand star predation by *A. americanus* could negatively impact sea scallop distributions (presence/absence) where sand star density exceeds 5600 individual km<sup>-2</sup> in the Hudson Shelf valley and its surrounding shelf edge within the MAB (Hart, 2006; Shank *et al.*, 2012).

Environmental drivers explaining patterns of observed SDM prediction error identified by validating SDM presence-absence predictions against commercial stock (presence-absences) within the commercial scallop beds considered here (see “Model fitting procedure and evaluation”) were determined by Multinomial Logistic Regression (MLR). MLR is a statistical classification method that generalizes logistic regression to multiclass problems, with more than two possible discrete outcomes, including (i) overprediction, (ii) underprediction, or (iii) correctly predicting known occurrences. Probabilities of each outcome were predicted from the tested environmental drivers of distribution and subsequently plotted in the same georeferenced (spatial) context to visualize and examine how these environmental drivers might influence patterns of prediction error.

## Results

### Accounting for SDM prediction error across genetic gradients?

#### Accuracy assessment during model training and testing

The RW species SDM (scenario 1) exhibited low overall predictive utility with a TSS = 0.53, the lowest of all tested scenarios (Table 1). Lineage level SDM based on the primary cline (scenario 2) exhibited moderate overall predictive utility (north TSS = 0.67 and south TSS = 0.7) during model training and testing (Table 1). Further gains in SDM accuracy north of the primary cline during model training and testing were evident after identifying two northern lineages under a secondary cline as in scenario 3, where North<sup>GSL</sup> lineage TSS = 0.80, and North<sup>NL</sup> lineage TSS = 0.86, both TSS scores of which equates to very good predictive utility (Table 1).

#### Validation against the distribution of commercial stocks that fall within the extent of each tested scenario

North of the primary cline, the RW species SDM (scenario 1) correctly classified only 37% of observed presences and absences (TSS = 0.37) in the SGS scallop stock north of the primary cline (Figure 2, Supplementary Tables S1 and S2). The northern lineage SDM (scenario 2) accurately classified 68% of observed presence and absence data (TSS = 0.68) in the SGS scallop stock. Finally, model performance for the SGS stock was optimal for the

**Table 1.** Accuracy of three SDM scenarios, including 1. No cline (RW species), 2. Primary cline (south vs. north lineages), and 3. Secondary cline (North<sup>GSL</sup> vs. North<sup>NL</sup> lineages).

Scenario	Model performance					
	TSS	Sensitivity	Specificity	Under	Over	Total
1. No cline						
RW	0.53	0.93	0.6	0.07	0.40	0.47
2. Primary cline						
South lineage	0.67	0.89	0.79	0.11	0.21	0.32
North lineage	0.7	0.87	0.84	0.13	0.16	0.29
3. Secondary cline (north)						
North <sup>GSL</sup> lineage	0.80	0.90	0.90	0.10	0.10	0.20
North <sup>NL</sup> lineage	0.86	0.90	0.95	0.10	0.05	0.15
Average	0.83	0.90	0.93	0.02	0.10	0.12

Model predictions are compared against known sea scallop (*P. magellanicus*) occurrences. SDM accuracy was assessed using a TSS, sensitivity, specificity, and related rates of under, over, and total prediction error.

North<sup>GSL</sup> lineage SDM (scenario 3), which accurately predicted 86% of observed presence and absence (TSS = 0.86) (Figure 2, Supplementary Tables S1 and S2). In other words, the high prediction error observed in the RW species SDM at the northern margin of the species distribution was considerably reduced when genetic structure was used to optimize the SDM domain incorporated as in scenario 3. Interestingly, >89% of observed presence and absences (TSS ≥ 0.89) documented in stocks south of the primary cline (MAB, GEO, BB, GB, SI, SPB; Figure 2 and Supplementary Figure S1b and c, and Supplementary Tables S1 and S2) were accurately predicted by both RW and lineage-specific SDMs (see southern lineage SDM under scenario 2), which equates to very high predictive utility regardless of whether or not genetic structure was incorporated in SDM.

#### Environmental drivers of distribution and predicted tolerances in relation to patterns in prediction error

The relative importance of environmental drivers (expressed as %) and associated thresholds for occurrence differed across SDM scenarios (Table 2). Summer (26%) and fall (38%) temperature climatology's, but not salinity, were important drivers of scallop distribution in RW species SDM (scenario 1) (Table 2). Summer (22%), fall (23%), and particularly winter (29%) temperature climatology's were important drivers of scallop distribution for stocks south of the primary cline (i.e. southern lineage, scenario 2 SDM) (Table 2). Winter temperature (47%) and salinity climatology's (29%) were most significant drivers in the northern lineage, scenario 2 SDM (Table 2). Winter salinity (53%) was an even more significant driver for scallop distributions north of the primary cline alongside winter temperature (26%) in the North<sup>GSL</sup> lineage SDM (scenario 3) (Table 2). In contrast, summer, fall, and winter temperature climatology's were the most important drivers of scallop distribution within the North<sup>NL</sup> lineage SDM in scenario 3 (Table 2).

Based on SDM predicted threshold for occurrence in Table 2, northern scallop lineages identified (scenarios 2 and 3) were predicted to be relatively more cold tolerant (by up to 5°C) than southern lineage in scenario 2, particularly during fall and winter conditions. A comparison of North<sup>GSL</sup> and North<sup>NL</sup> lineages along the secondary cline (scenario 3) revealed important

differences in salinity tolerance (Table 2), with a lower tolerance limit (23–24 PSU) than all other lineage scenarios, whose lower tolerance limits varied from 26 to 32 PSU (Table 2).

Models incorporating genetic structure, driven by localized temperature and salinity tolerances described above, more clearly differentiated suitable vs. unsuitable habitat north of the primary cline (Figure 1b–d). Not accounting for genetic structure in SDMs (scenario 1) resulted in clusters of prediction error, particularly underprediction error, throughout the SGS stock within the GSL region (Figure 3a). In contrast, underprediction error was more limited to the Northumberland Strait (within the GSL region) in SDM that included the northern genetic lineage stemming from the primary cline (scenario 2); that in turn is predicted to be relatively more cold tolerant than lineages in SDM without genetic structure (Figure 3b). Relative to the rest of the GSL region, the Northumberland Strait is characterized by lower salinity. Inclusion of the North<sup>GSL</sup> lineage in SDM (scenario 3), which is predicted to be more tolerant of colder temperatures and/or lower salinity than the other tested SDM scenarios, resulted in the most accurate depiction of the SGS stock within the GSL region (Table 2, Figure 3c).

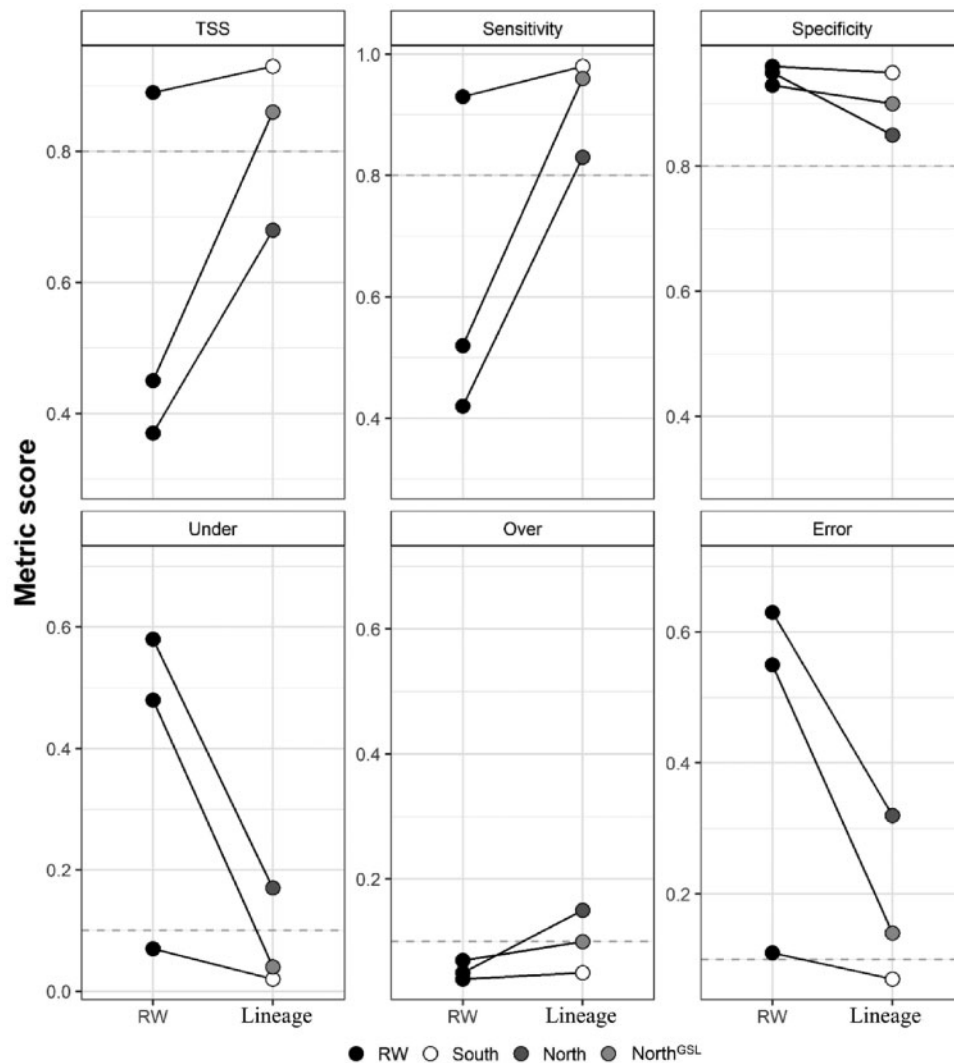
### Accounting for SDM prediction error across abiotic and biotic interaction gradients

#### Abiotic gradients

Spatial clustering of SDM prediction error is indicative of areas where important environmental drivers of distribution have been poorly characterized. MLR analyses revealed that BT and topographic complexity, expressed as curvature or slope, were strongly correlated with clusters of over or underprediction error in and around shelf edges and channels associated with MAB and GEO scallop stocks (Supplementary Table S3, Figures 4 and 5).

In the MAB, the probability of under or over predicting sea scallop distributions increased at summer temperatures between 10 and 18°C (Supplementary Figure S2a). A large proportion of over- and underprediction errors were distributed around the upper summer temperature threshold for establishment (14°C) (Supplementary Figure S2a), while the probability of under predicting sea scallop distribution increased at the steepest slopes (Supplementary Figure S2b). In GEO, the probability of correctly predicting sea scallop distributions decreased with progressively more convex bottom type (depressions), or at summer temperatures between 10 and 12°C and winter temperatures between 5 and 6.4°C (Supplementary Figure S3a–c). Under, but not overprediction error, within GEO was associated with the transition from flat to more convex bottom types (Supplementary Figure S3c), while overprediction, and to a much lesser extent, underprediction error, was positively associated with summer and negatively associated with winter BT (Supplementary Figure S3a and b).

In the MAB, underprediction errors were clustered around and strongly associated with abrupt transitions in bottom slope in areas that encompassed channels and canyons within the MAB, and the continental shelf edge (Figure 4a). In these channels, canyons, and the shelf edge, summer BT was also strongly associated with clusters of underprediction error, especially where summer BT was close to and exceeded the upper summer temperature threshold for establishment (~14°C) (Figure 4b, Supplementary Figure S2a). Underprediction errors were also strongly associated with areas within GEO where the bottom was



**Figure 2.** Validation of SDM predictions against commercial stocks that fall within the geographic extent of each of three tested lineage scenarios based on predicted vs. observed presences and absences. To facilitate bivariate comparisons with lineage level SDM, predictions from scenario 1. No cline (RW species SDM) was partitioned into the corresponding domains of the lineage level SDM predictions based on scenario 2. Primary cline (south or north lineage), and scenario 3. Secondary cline (North<sup>GSL</sup> lineage). Model performance, assessed with the TSS, sensitivity, specificity, and related rates of under, over, and total prediction error, is plotted as a bivariate relationship for. RW species SDM vs. south lineage SDM predictions were validated against commercial stocks south of the primary cline (MAB + GEO + BB + GB + SI + SPB). While, RW SDM vs. north lineage SDM predictions were validated against commercial stocks north of the primary cline (SGS stock). Finally, RW SDM vs. North<sup>GSL</sup> lineage SDM predictions were validated against stocks north of the primary cline (GF stock).

progressively more convex or channelled; areas that reflect abrupt transitions in depth at various points along the shelf edge and within several channels in the GEO bank (Figure 5a). Within these areas a large proportion of overprediction errors were also distributed and associated with both the upper summer temperature threshold for establishment ( $\sim 14^{\circ}\text{C}$ ) and the lower winter temperature threshold for establishment ( $4.6^{\circ}\text{C}$ ) (Figure 5b and c, Supplementary Figure 3a and b).

#### Biotic interaction gradients

Overprediction error was closely associated with areas where sea stars exceeded  $5600$  individual  $\text{km}^{-2}$ , typically within deeper waters around shelf edges and within the Hudson Shelf Valley in the

MAB (Figure 6). Here, overprediction errors were strongly associated with increasing sea star density (Figure 6).

#### Discussion

Empirical sea scallop distributions were most accurately predicted with lineage-specific SDM when compared to models based on the RW species distribution. Contiguous omissions errors observed in northern scallop distributions were reduced with lineage-specific SDM, particularly when fine-scale North<sup>GSL</sup> and North<sup>NL</sup> genetic structure (modelling scenario 3) was incorporated. Consequently, this study supports growing evidence in terrestrial environments (Benito Garzón et al., 2011; Bocedi et al., 2013; Oney et al., 2013; Homburg et al., 2014; Romero et al., 2014; Valladares et al., 2014) that genetic lineages are also the

**Table 2.** Characterization of how seasonal temperature and salinity climatologies shape the predicted distribution of sea scallop (*P. magellanicus*) within each SDM scenario as defined by percentage contribution (%) of BT and bottom salinity to model performance, lower thresholds (Tr.) for detection based on seasonal temperature and salinity tolerances.

Seasonal drivers		Scenario of lineage structure										
		1. No cline		2. Primary cline				3. Secondary cline (north)				
		RW		South lineage		North lineage		North <sup>GSL</sup> lineage		North <sup>NL</sup> lineage		
		%	Tr.	%	Tr.	%	Tr.	%	Tr.	%	Tr.	
Spring	PSU	6	28	5	31	4	26	4	23	6	26	
	°C	6	3	11	4	8	-1	1	0	0	-2	
Summer	PSU	6	28	0	32	4	27	7	23	0	27	
	°C	26	4	22	6	8	2	2	0	36	2	
Fall	PSU	3	27	2	31	7	27	3	24	0	NA	
	°C	38	5	23	5	2	-1	2	1	25	-1	
Winter	PSU	9	27	6	32	29	27	53	24	4	29	
	°C	4	1	29	3	47	-1	26	-1	29	-2	

Tested scenarios included 1. No cline (RW species), 2. Primary cline (south vs. north lineages), and 3. Secondary cline (North<sup>GSL</sup> vs. North<sup>NL</sup> lineages).

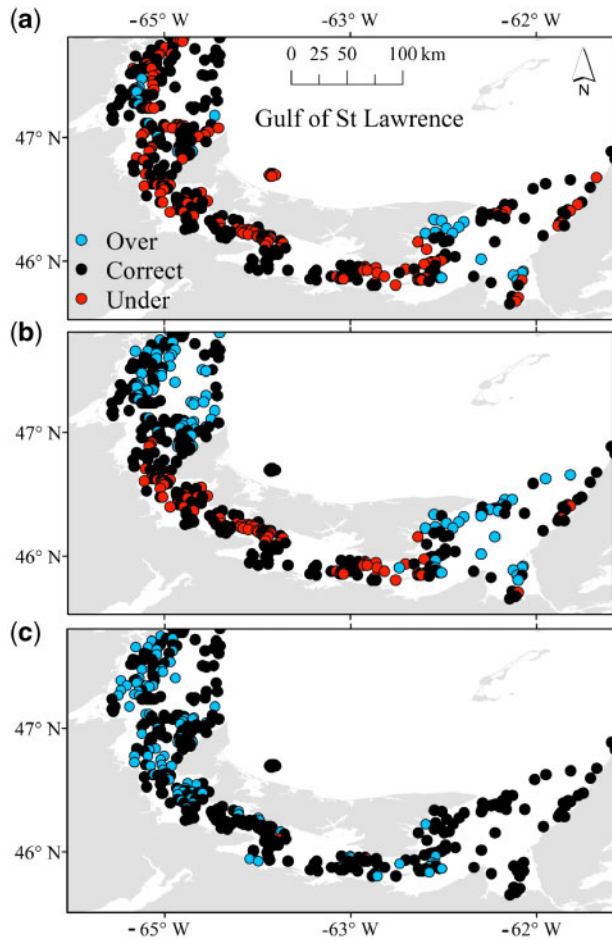
most appropriate biological unit to base individual SDM for marine species. Species can be highly heterogeneous entities as a result of environmentally driven demographic and adaptive processes (Hernandez *et al.*, 2006). RW species SDM calibrated with occurrence data across a species range effectively treat a species as a single biological entity, where all lineages within a species would have the same average trait values and the same reactions norms (magnitude of phenotypic plasticity) such that their fundamental niches are the same and equal to the overall fundamental niche of the species (Schwartz, 2012; Valladares *et al.*, 2014). Local specialization, whereby disjunct lineages exhibit local differences in the range of their climatic (physiological) tolerances across their distribution violates this assumption of ecological uniformity. Ecologically distinct lineages with different environmental responses and tolerances, and consequently different habitat requirements, may not be accurately represented in species RW SDM (Supplementary Figure S4). RW species SDM can smooth over important differences in suitable habitat among genetically distinct groups to create a uniform response (e.g. occurrence probability) to the observed range of environmental conditions; rather than the multiple lineage-specific responses required if lineages are ecologically distinct, as in this study (Supplementary Figure S4).

Northern lineages (North<sup>NL</sup> and North<sup>GSL</sup>) for example were predicted to be more cold tolerant (by up to 5°C) than the southern lineage or when the species was modelled as a single panmictic unit (species RW SDM) (Figure 4a). The North<sup>GSL</sup> lineage (scenario 3) was also predicted to be more tolerant of lower salinity's than the other lineages and SDM scenarios (Figure 4a). Large clusters of occurrences excluded by RW species SDM in colder or lower salinity environments in the northern margins of the sea scallops range were in turn accurately predicted in lineage level SDM (based on one or two northern lineages). In effect, lineage level SDM identified the same suite of environmental variable tolerances (i.e. colder temperatures and lower salinity in the Northern range margins) that account for observed geographic variation in genetic structure (Lehnert *et al.*, 2018; Stanley *et al.*, 2018; Van Wyngaarden *et al.*, 2018); with one important caveat. Despite finding evidence of local specialization in the northern margins of the sea scallops range we did not find any evidence of

local specialization further south. RW species and southern lineage SDM performed equally well and predicted a similar suite of environmental tolerances, including tolerance of warmer temperatures and a preference for more stable and higher salinity than northern lineage SDM. An outstanding question then is why scallop at the northern but not southern range margins might exhibit local specialization that cannot be accounted for in RW species SDM?

The answer to this question is dependent on defining where core and margin populations actually occur and then explaining why local specialization might not be selected for across all environmental margins of a species range. At species range margins, distributional limits are typically imposed by environmental constraints and (i) individual fitness is generally lower than in the centre (core) of its distribution, while (ii) selection pressure for traits that enhance fitness are stronger than in areas where lineages experience more favourable conditions (Valladares *et al.*, 2014). As such, it has been predicted that core vs. margin lineages will differ significantly in trait means and plasticity patterns that in turn influence genetic structure and associated physiological response curves (Valladares *et al.*, 2014). SDM often under-predict species distributions at the margins of their range (Franklin, 2010); especially if they are broadly distributed (Hernandez *et al.*, 2006), and span steep environmental gradients (Watling *et al.*, 2015), as in this study. The sea scallop has split into spatially and ecologically disjunct genetic groups that can tolerate either; (i) colder temperatures and lower salinity in the northern margins of its range (represented by the northern lineages), or (ii) warmer temperatures and stable high salinity's in a much larger area subsuming the southern lineage in the south of its range that effectively represents the core of sea scallops current distribution. Given that the core and northern margin of the sea scallops range have been described, the question remains as to why there appears to be no evidence of ecological or genetic differentiation in the southern range margins?

Northern lineages constitute a much smaller proportion of the overall scallop population both in terms of distribution of occurrences, and density compared to the southern lineage. Northern lineages are also subjected to much lower salinities (North<sup>GSL</sup>) and extremes of cold (North<sup>GSL</sup> and North<sup>NL</sup>) than sea scallop in



**Figure 3.** Distribution of overprediction (blue dots) and underprediction (red dots) error within the GSL for corresponding SDM scenarios: (a) No cline (RW species), (b) Primary cline (north lineage), and (c) Secondary cline (North<sup>GSL</sup> lineage).

the southern lineage. The southern lineage on the other hand occupies a much larger area than the scattered stocks within the northern lineages and is composed of many large and relatively well-connected (in terms of exchange of individuals) stocks that are predicted to be more tolerant of warm extremes than those in the north. Therefore, it follows that ecological response curves based on species RW SDMs would not capture local environmental requirements of lineages in the northern range margins that occupy a relatively small portion of the environmental domain.

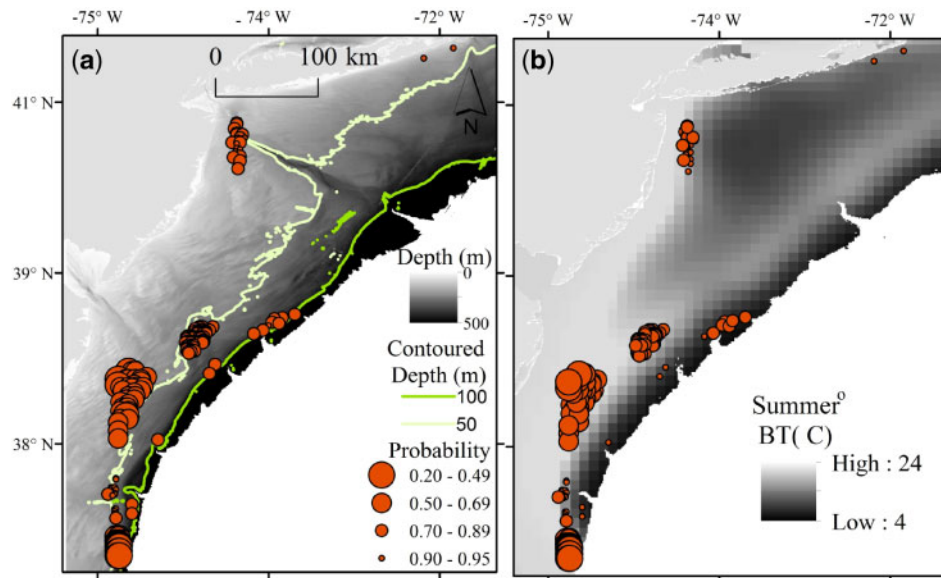
Sea scallop at the southern range margins also experience warmer temperatures that are largely unsuitable for scallop (Davies *et al.*, 2014; Munroe *et al.*, 2018), yet both RW species and lineage level SDM captured distributions equally well at the southern range margins that are driven by extremes of warmer temperatures. Why sea scallop distributions were accurately captured at the southern range margins but not the northern range margins in RW species SDM could be attributed to differences in their relative connectivity in terms of exchange of individuals from favourable to unfavourable locations. Put simply, scallop within the much larger southern lineage are likely better connected than scallop in the more spatially isolated northern lineages, as suggested by fine-scale patterns in genetic structure at

northern latitudes (Lehnert *et al.*, 2018). Individuals dispersing from more favourable habitat can maintain stocks that persist in more unfavourable habitat (e.g. extremes of warmer temperatures) in the southern margins (Munroe *et al.*, 2018), but not the northern margins where stocks are largely disconnected from each other. A discrepancy in connectivity patterns among northern and southern range margins in turn would help to explain why lineage level SDM more accurately captured scallop distributions in the north than species RW SDM, and why species level and lineage level SDM performed similarly in the south, which effectively covers the core of the scallops distribution and where larger and better connected stocks than in the north are evident.

Local specialization and associated patterns in genetic structure (Lehnert *et al.*, 2018), in the northern but not the southern range margins of sea scallop could also be explained, in part, by the climatic variability hypothesis (CVH). CVH has been addressed and confirmed in different groups of terrestrial animals (reviewed by Valladares *et al.*, 2014), but has yet to be demonstrated in the marine realm. In essence, CVH states that, as the range of climatic fluctuation from warm to cold experienced by terrestrial organisms throughout a year increases with latitude, individuals at higher latitudes should have greater physiological flexibilities to cope with extremes of cold, yet should retain some of their tolerance to warmer conditions (Janzen, 1967). Indeed, niche evolution can effectively be “frozen” in the warmer, but not the colder, extremes of a species range (Araújo *et al.*, 2013). Nonetheless, the extent to which the range of climatic fluctuation from warm to cold increases with latitude in tandem with physiological flexibilities in sea scallop has yet to be quantified in the NW Atlantic.

In summary, employing genetic structure to define the number and boundaries of SDMs significantly improved model performance (Table 2). Yet, clusters of prediction errors persisted in lineage-specific SDMs. Additional SDM prediction errors tended to cluster in grids that were associated with aspects of topographic complexity (i.e. bottom slope and curvature) and areas of short spatial transitions in seasonal BT. Topography and BT can both influence species distributions as the shelf break encompasses regions of the ocean bottom with steep depth gradients (high slope) and is often associated with abrupt and predictable changes in temperature. The 8-km grid resolution employed in this study may span a bottom depth of 20–300 m with a corresponding temperature change of up to 8°C near the shelf break. The average temperatures per grid cell along these shelf edges and channels consequently reflect very large changes in depth (up to 300 m). Consequently, the weighted averages across grid-cells and the high variance around BT estimates are not representative of shallow, localized scallop habitat. Such extensive spatial variability, including abrupt spatial shifts in temperature, can hinder SDM performance (Valavanis *et al.*, 2008) when patterns in prediction error are the result of poor characterization (i.e. resolution) of limiting variables (e.g. temperature) (Lowen *et al.*, 2016).

For example, prediction error attributable to temperature in this study likely stems from interpreting gridded temperatures in and around shelf breaks and channels as either suitable or unsuitable for establishment (based on predicted thresholds) when the opposite holds. Grid cells on the shelf break can span shallow (warm) to very deep (colder) waters and weighted averages would therefore be colder than the warmer waters found at the shallowest depths that are most suitable for scallops. Interestingly, at the southern tip of the MAB the opposite is true. Here, more suitable temperatures for scallop are found in deeper channels that



**Figure 4.** Probability of correctly predicting sea scallop (*P. magellanicus*) distributions in the MAB. Larger circles identify areas where progressively higher prediction error is more likely driven by (a) bottom slope or (b) BT in areas of SDM underprediction error. Probability intervals and size of circles are based on the Jenks natural breaks classification method. Probability scaling shown in panel a also applies to panel b. Depth contours (50 and 100 m) that highlight breaks within the shaded bathymetry (m), and benthic summer temperature are shown for reference in panels a and b, respectively.

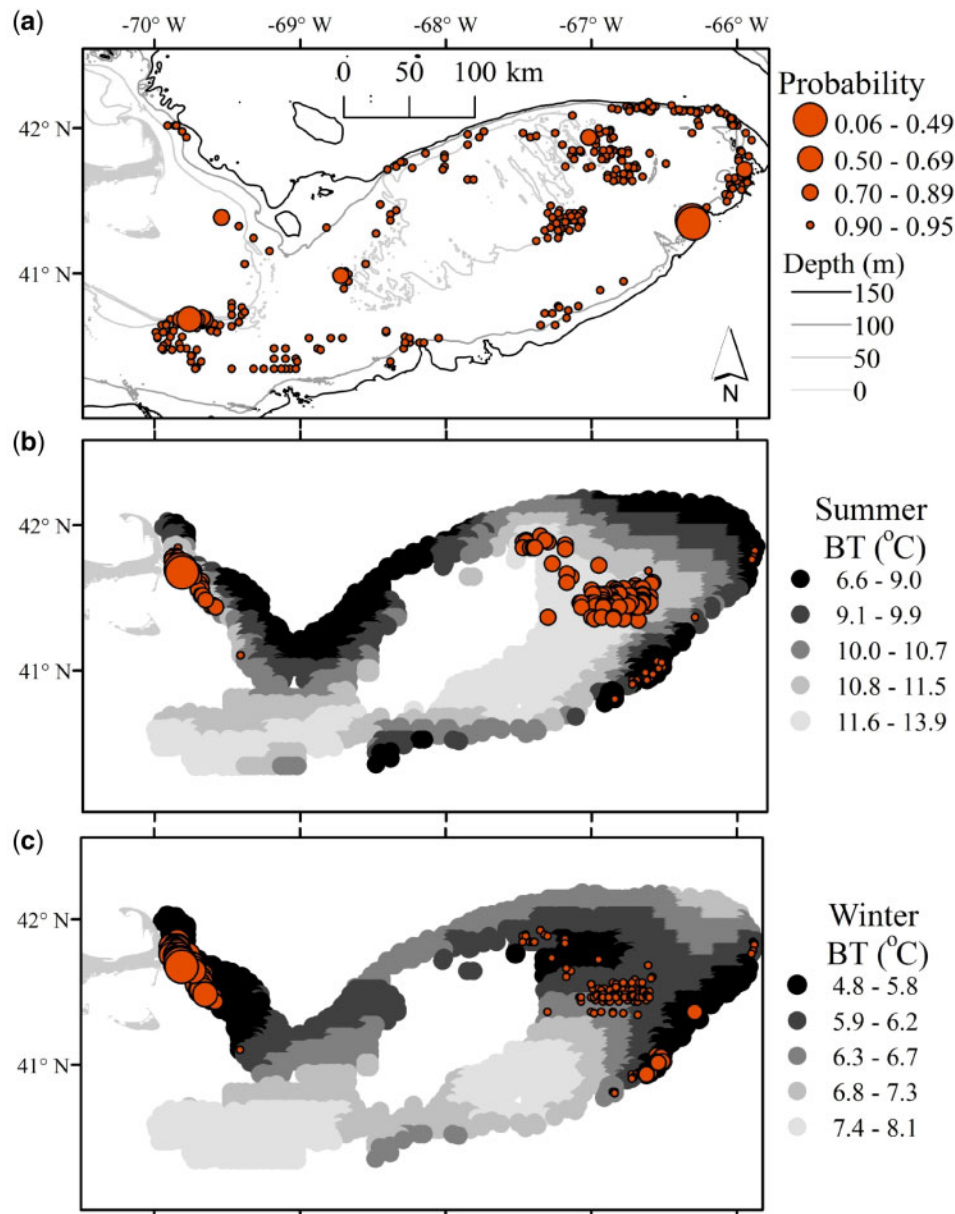
punctuate bodies of extremely warm water around shallow ledges that are largely unsuitable for scallop. Both these scenarios result in underprediction error. In other words, the gridded environmental data likely “smoothed” over important variations in temperature in and around shelf breaks and channels that limit scallop establishment there. In our case limiting temperatures equated to  $>14^{\circ}\text{C}$  in the summer, or  $<5^{\circ}\text{C}$  in the winter and were in close agreement with reported observations (Hart and Chute, 2004). It follows that modelled grids were too coarse in regions of our domain to resolve limiting environmental variation (here temperature), a problem observed in SDM for many species in both marine and freshwater environments (Davies and Guinotte, 2011; Sbrocco and Barber, 2013; Basher et al., 2014). Consequently, the spatial resolution of environmental gridded temperature should be increased sufficiently to match spatial patterns of limiting temperature found within shelf breaks and channels.

Biotic interactions including competition, predation, and dispersal may also confound SDM predictions (Dormann, 2007). For example, pre-recruit predation by the sand star (*A. americanus*) in the MAB have been hypothesized to constrain the distribution of sea scallop (Hart, 2006; Shank et al., 2012). Not taking the spatial distribution of predation into account could lead to overprediction error where scallops are predicted to be present, but are not due to predation. The presence of sand stars was not associated with SDM overprediction error across the extent of the sea scallops range in the MAB. Spatial patterns in SDM prediction error were instead driven by a density-dependent predator–prey relationship. In deeper waters around shelf breaks in the MAB and within the Hudson Shelf valley overprediction error was driven by the distribution of the sand star once it exceeded  $\sim 1\text{ m}^{-2}$  as reported by Hart (2006) and Shank et al., (2012).

Overall, the genetic, abiotic, and biotic–environmental gradients examined here are commonly invoked to shape marine species distributions and must be appropriately characterized in SDM to prevent clusters of prediction error. Local specialization driven by sharp temperature gradients that align with distinct genetic lineages may be pervasive in the North West Atlantic across a range of taxa with very different life-histories (Stanley et al., 2018). Temperature and salinity gradients have also been shown to drive patterns of genetic structure in other bioregions that encompass Atlantic herring (*Clupea harengus*) (Limborg et al., 2012), Atlantic cod (*Gadus morhua*) (Berg et al., 2015), purple sea urchin (*Strongylocentrotus purpuratus*) (Pespeni and Palumbi, 2013), European hake (*Merluccius merluccius*) (Milano et al., 2014), and Chinook salmon (*Oncorhynchus tshawytscha*) (Hecht et al., 2015). It follows that environmentally associated lineages in the marine environment should be treated as the biological unit of interest in marine SDM. Furthermore, identifying and correlating patterns of prediction error across limiting climatological and biotic gradients, as in this study, also encourages development of more appropriate environmental predictors, namely a spatially explicit and higher-resolution characterization of limiting environmental variables (Basher et al., 2014; Lowen et al., 2016). Reliable predictions of suitable habitat in marine environments are in turn required to address emerging ecological and evolutionary problems including:

#### *Management of species responses under changing climates:*

Targeted conservation measures to protect species-at-risk or adaptive management of commercially important stocks based on SDM cannot be achieved without accurate predictions of potential distribution coupled to an understanding of environmental drivers that shape species distributions under current and future climate scenarios (Franklin, 2010). Using sea scallop as a case study species, we have shown that large clusters of omission error at the northern margins of a species range can be accurately

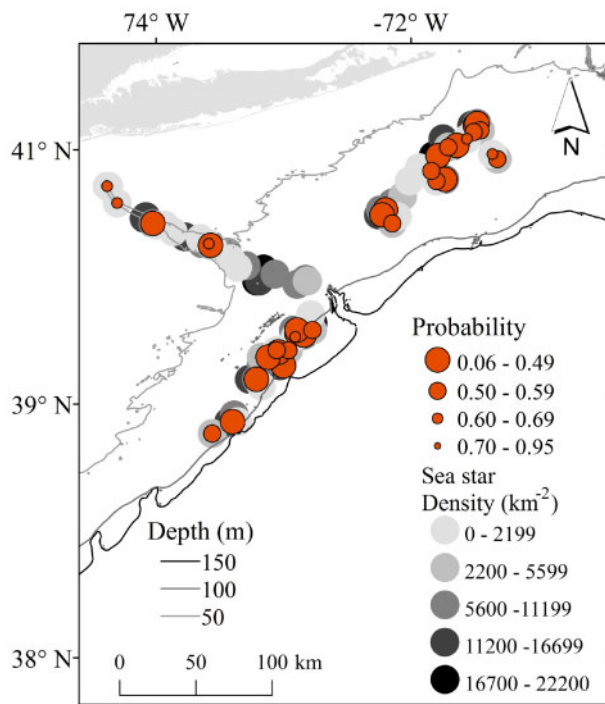


**Figure 5.** Probability of correctly predicting sea scallop (*P. magellanicus*) distributions on GEO. Larger circles identify areas where prediction error is more likely driven by (a) bottom curvature and (b) summer or (c) winter BT. Probability intervals and size of circles in areas of (a) SDM underprediction, or (b and c) SDM overprediction error are based on the Jenks natural breaks classification method. In (a) Contoured depth (m) or benthic (b) summer or (c) winter temperature are also shown. Probability scaling shown in panel a also applies to panels b and c.

resolved after treating lineages separately in SDM. Higher resolution climatology and consideration of biotic interactions can help to better resolve clusters of error that manifest in areas where climatic and biotic gradients are poorly resolved around oceanographic fronts that drive important and fine-scale patterns in species distributions. The issue of not resolving clusters of prediction error in SDM at the margins of species range is particularly acute when predictions of a species distribution are extrapolated across climate scenarios to infer range shifts under climate change (Valladares *et al.*, 2014). In other words, northward shifts in sea scallop species distributions in our study across climate scenarios would not be accurately resolved if the environmental limits of a species range are poorly characterized in SDM.

#### *Appropriate modelling of invasive species:*

Anthropogenic dispersal among suitable habitat has greatly extended the reach of many of the world's marine species beyond their native range resulting in the proliferation of invasive species across the world's oceans (Franklin, 2010). Species RW SDM may span very large extents to account for this global exchange of AIS (Franklin, 2010), which in turn carries the risk of treating disjunct lineages that have distinct environmental tolerances as one ecological entity. The resultant SDM would have limited utility when predicting potential distributions in local areas of interest (e.g. for risk assessments) due to averaging physiological tolerances from disjunct lineages when they should be treated separately. If the spread of genetically and ecologically distinct lineages of



**Figure 6.** Probability of correctly predicting sea scallop (*P. magellanicus*) distributions on the Hudson Shelf Valley and surrounding shelf edge. Larger circles identify areas where SDM overprediction error is more likely driven by the sand star (*A. americanus*). Probability intervals and size of circles are based on Jenks natural breaks classification method. Sea star density is also shown for reference (grey-scaled circles).

marine invasive species are to be appropriately managed and understood they should be carefully considered in SDM that subsume large distributional ranges or that project suitable habitat to distant locations that have yet to be invaded. European green crab (*Carcinus maenas*) is one of the most problematic invasive species in marine environments and are currently present on every continent with the exception of Antarctica (Roman, 2006). Green crab in Atlantic Canada has been treated as a single species in SDM-based risk assessment. However, European green crab (*C. maenas*) in the NW Atlantic consist of two distinct lineages that originated from northern and southern locations in its native range (Jeffery *et al.*, 2018). The northern ecotype is more tolerant of colder environments while the southern ecotype is more tolerant of warmer waters (Jeffery *et al.*, 2018). Consequently, the two ecotypes should be treated separately in SDM that are used to predict their distributions in the NW Atlantic.

## Conclusion

SDMs are increasingly being used to predict and understand changes in marine species distributions and this study highlights how the accuracy of marine SDM can be dependent on capturing the appropriate biological unit for modelling (e.g. ecologically distinct genetic lineages stemming from genetic gradients as in this study) and the importance of adequately identifying limiting abiotic and biotic gradients that can drive clusters of prediction error. First-order approximations of sources of prediction error in SDM can serve as a basis for refined investigations requiring more time and resources.

## Supplementary data

Supplementary material is available at the ICES/JMS online version of the manuscript.

## Acknowledgements

We thank the scallop stock assessment division (in particular David Keith and Leslie Nasmith) at the Bedford Institute of Oceanography and Devorah R. Hart (National Marine Fisheries Service, USA) for providing sea scallop distributional data (presence-absence locations) for commercially important off-shore scallop stocks. We would also like to thank Xeilang Wang for providing monthly data (BT and salinity) from the Global Ocean Reanalyses and Simulations (GLORYS2V2) which have been adapted to the North West Atlantic by Fisheries and Oceans Canada (Xeilang Wang). We also thank Chris Legault (National Marine Fisheries Service, USA) for valuable feedback on the manuscript.

## Funding

This research was supported by funding from Fisheries and Oceans Canada's Strategic Program for Ecosystem-based Research and Advice (SPERA), and the Government of Canada's Genomic Research Development Initiative (GRDI).

## References

- Allouche, O., Tsoar, A., and Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43: 1223–1232.
- Araújo, M. B., Ferri-Yañez, F., Bozinovic, F., Marquet, P. A., Valladares, F., and Chown, S. L. 2013. Heat freezes niche evolution. *Ecology Letters*, 9: 1206–1219.
- Basher, Z., Bowden, D. A., and Costello, M. J. 2014. Diversity and distribution of deep-area shrimps in the Ross Sea region of Antarctica. *PLoS One*, 9: e103195.
- Beazley, L., Lirette, C., Sabaniel, J., Wang, Z., Knudby, A., and Kenchington, E. 2016. Characteristics of environmental data layers for use in species distribution modelling in the Gulf of St. Lawrence. Canadian Technical Report of Fisheries and Aquatic Sciences, 3154: viii + 357.
- Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., Jakobsen, K. S. *et al.* 2015. Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution*, 7: 1644–1663.
- Bocedi, G., Atkins, K. E., Liao, J., Henry, R. C., Travis, J. M., and Hellmann, J. J. 2013. Effects of local adaptation and interspecific competition on species' responses to climate change. *Annals of the New York Academy of Sciences*, 1297: 83–97.
- Brown, J. L. 2014. SDMtoolbox: a Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, 5: 694–700.
- Chase, J. M., and Leibold, M. A. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press, Chicago.
- Cromwell, T., and Reid, J. L. 1956. A study of oceanic fronts. *Tellus*, 8: 94–101.
- Davis, M. B., and Shaw, R. G. 2001. Range shifts and adaptive responses to quaternary climatechange. *Science*, 292: 673–679.
- Davies, A. J., and Guinotte, J. M. 2011. Global habitat suitability for framework-forming cold-watercorals. *PLoS One*, 6: e18483.
- Davies, K. T., Gentleman, W. C., DiBacco, C., and Johnson, C. L. 2014. Semi-annual spawning in marine scallops strengthens larval recruitment and connectivity on Georges Bank: a model study. *Marine Ecology Progress Series*, 516: 209–227.

- DFO. 2016. Commercial Fisheries: Landings. Fisheries and Oceans Canada. Canadian Manuscript Report of Fisheries and Aquatic Sciences.
- Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, 8: 387–397.
- Fitzpatrick, M. C., and Keller, S. R. 2015. Ecological genomics meets community-level modeling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18: 1–16.
- Franklin, J. 2010. *Mapping Species Distributions. Spatial Inference and Prediction*. Cambridge University Press, Cambridge.
- Galarza, J. A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G. F., and Rico, C. 2009. The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America*, 106: 1473–1478.
- Garzón, B., Alía, R., Robson, T. M., and Zavala, M. A. 2011. Intra-specific variability and plasticity influence potential tree species distributions under climate change. *Global Ecology and Biogeography*, 20: 766–778.
- Hart, D. 2006. Effects of sea stars and crabs on sea scallop *Placopecten magellanicus* recruitment in the Mid-Atlantic Bight (USA). *Marine Ecology Progress Series*, 306: 209–221.
- Hart, D. R., and Chute, A. S. 2004. Essential Fish Habitat Source Document: Sea Scallop, *Placopecten magellanicus*, Life History and Habitat Characteristics, 2nd edn. NOAA Tech Memo, NMFS NE-189. Northeast Fisheries Science Center, Woods Hole, MA.
- Hecht, B. C., Matala, A. P., Hess, J. E., and Narum, S. R. 2015. Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology*, 24: 5573–5595.
- Hernandez, P. A., Graham, C. H., Master, L. L., and Albert, D. L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29: 773–785.
- Homburg, K., Brandt, P., Drees, C., and Assmann, T. 2014. Evolutionarily significant units in a flightless ground beetle show different climate niches and high extinction risk due to climate change. *Journal of Insect Conservation*, 18: 781–790.
- Janzen, D. H. 1967. Why mountain passes are higher in the tropics. *American Naturalist*, 101: 233–249.
- Jeffery, N. W., Bradbury, I. R., Stanley, R. E., Wringe, B. F., Van Wyngaarden, M., Lowen, J. B., McKenzie, C. H. *et al.* 2018. Genome-wide evidence of environmentally mediated secondary contact of European green crab (*Carcinus maenas*) lineages in eastern North America. *Evolutionary Applications*, 11: 869–882.
- Lehnert, S. J., DiBacco, C., Van Wyngaarden, M., Jeffery, N. W., Lowen, J. B., Sylvester, E. V. A., Wringe, B. F. *et al.* 2018. Fine-scale genetic structure between inshore and offshore lineages of sea scallop (*Placopecten magellanicus*) associated with temperature. In revision: *Heredity*, 10: 69–80.
- Limborg, M. T., Helyar, S. J., De Bruyn, M., Taylor, M. I., Nielsen, E. E., Ogden, R. O. B., Carvalho, G. R. *et al.* 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, 21: 3686–3703.
- Lobo, J. M., Jiménez-Valverde, A., and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17: 145–151.
- Lowen, J. B., and DiBacco, C. 2017. Distributional changes in a guild of non-indigenous tunicates in the NW Atlantic under high resolution climate projections. *Marine Ecology Progress Series*, 570: 173–186.
- Lowen, J., McKindsey, C., Therriault, T., and DiBacco, C. 2016. Effects of spatial resolution on predicting the distribution of aquatic invasive species in nearshore marine environments. *Marine Ecology Progress Series*, 556: 17–30.
- Mateo, R. G., Croat, T. B., Felicísimo, Á. M., and Muñoz, J. 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distributions*, 16: 84–94.
- May, R. M., and MacArthur, R. H. 1972. Niche overlap as a function of environmental variability. *Proceedings of the National Academy of Sciences of the United States of America*, 69: 1109–1113.
- Merow, C., Smith, M. J., and Silander, J. A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36: 1058–1069.
- Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M. *et al.* 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23: 118–135.
- Munroe, D. M., Haidvogel, D., Caracappa, J. C., Klinck, J. M., Powell, E. N., Hofmann, E. E., Shank, B. V. *et al.* 2018. Modeling larval dispersal and connectivity for Atlantic sea scallop (*Placopecten magellanicus*) in the Middle Atlantic Bight. *Fisheries Research*, 208: 7–15.
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., and Anderson, R. P. 2014. ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for MaxEnt ecological niche models. *Methods in Ecology and Evolution*, 5: 1198–1205.
- NEFSC. 2014. Assessment of Atlantic sea scallops. In Report of the 50th Northeast Regional Stock Assessment Workshop (50th SAW). Stock Assessment Committee (SARC), pp 483–631. *Consensus Summary of Assessments NEFSC Reference Document*, 14-09, Northeast Fisheries Science Center.
- Oney, B., Reineking, B., O'Neill, G., and Kreyling, J. 2013. Intraspecific variation buffers projected climate change impacts on *Pinus contorta*. *Ecology and Evolution*, 3: 437–449.
- Pespeni, M. H., and Palumbi, S. R. 2013. Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology*, 22: 3580–3597.
- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., and Blair, M. E. 2017. Opening the black box: an open-source release of MaxEnt. *Ecography*, 40: 887–893.
- Phillips, S. J., Dudík, M., and Schapire, R. E. 2004. A maximum entropy approach to species distribution modeling. In Proc 21st Int Conf Machine Learning, pp. 655–662. ACM Press, New York, NY.
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A. *et al.* 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5: 397–406.
- Roman, J. 2006. Diluting the founder effect: cryptic invasions expand a marine invader's range. *Proceedings of the Royal Society B*, 273: 2453–2459.
- Romero, D., Olivero, J., Márquez, A. L., Báez, J. C., and Real, R. 2014. Uncertainty in distribution forecasts caused by taxonomic ambiguity under climate change scenarios: a case study with two new species in mainland Spain. *Journal of Biogeography*, 41: 111–121.
- Sbrocco, E. J., and Barber, P. H. 2013. MARSPEC: ocean climate layers for marine spatial ecology. *Ecology*, 94: 979.
- Schwartz, M. W. 2012. Using niche models with climate projections to inform conservation management decisions. *Biological Conservation*, 155: 149–156.
- Shank, B. V., Hart, D. R., and Friedland, K. D. 2012. Post-settlement predation by sea stars and crabs on the sea scallop in the

- Mid-Atlantic Bight. Marine Ecology Progress Series, 468: 161–177.
- Silva, A. G. D., Paula, A. F. D., Fleury, B. G., and Creed, J. C. 2014. Eleven years of range expansion of two invasive corals (*Tubastraea coccinea* and *Tubastraea tagusensis*) through the southwest Atlantic (Brazil). Estuarine, Coastal and Shelf Science, 141: 9–16.
- Smith, W. H. F., and Sandwell, D. T. 1997. Global seafloor topography from satellite altimetry and ship depth soundings. Science, v. 277: 1957–1962.
- Stanley, R. E., DiBacco, C., Lowen, J. B., Beiko, R. G., Jeffery, N. W., Van Wyngaarden, M., Bentzen, P. *et al.* 2018. A climate-associated multispecies cryptic cline in the northwest Atlantic. Science Advances, 4: eaaq0929.
- Therriault, T. W., and Herborg, L. M. 2008. Predicting the potential distribution of the vase 488 tunicate *Ciona intestinalis* in Canadian waters: informing a risk assessment. ICES, 65: 788–794.
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. Global Ecology and Biogeography, 21: 272–281.
- Valavanis, V. D., Pierce, G. J., Zuur, A. F., Palialexis, A., Saveliev, A., Katara, I., and Wang, J. J. 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. Hydrobiologia, 612: 5–20.
- Valladares, F., Matesanz, S., Guilhaumon, F., Aratijo, M. B., Balaguer, L., Benito-Garzon, M., Cornwell, W. *et al.* 2014. The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecology Letters, 17: 1–14.
- Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodriguez-Ezpeleta, N., Jeffery, N. W., and Bradbury, I. R. 2018. Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*. Ecology and Evolution, 8: 2824–2841.
- Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodriguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E. *et al.* 2017. Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. Evolutionary Applications, 10: 102–117.
- Watling, J. I., Brandt, L. A., Bucklin, D. N., Fujisaki, I., Mazzotti, F. J., and Románach, S. S. 2015. Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. Ecological Modelling, 309: 48–59.

Handling editor: Lorenz Hauser