

ECOGRAPHY

Research

Trade-offs in covariate selection for species distribution models: a methodological comparison

Stephanie Brodie, James T. Thorson, Gemma Carroll, Elliott L. Hazen, Steven Bograd, Melissa A. Haltuch, Kirstin K. Holsman, Stan Kotwicki, Jameal F. Samhouri, Ellen Willis-Norton and Rebecca L. Selden

S. Brodie (<https://orcid.org/0000-0003-0869-9939>) ✉ (sbrodie@ucsc.edu), G. Carroll, E. L. Hazen and S. Bograd, *Inst. of Marine Science, Univ. of California Santa Cruz, Santa Cruz, CA, USA, and Environmental Research Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Monterey, CA, USA.* – J. T. Thorson and K. K. Holsman, *Habitat and Ecosystem Process Research program, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA.* – M. A. Haltuch, *Fishery Research Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA.* – S. Kotwicki, *Resource Assessment and Conservation Engineering Division, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA.* – J. F. Samhouri, *Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA.* – E. Willis-Norton, *Dept of Ecology and Evolutionary Biology, Univ. of California Santa Cruz, Santa Cruz, CA, USA.* – R. L. Selden, *Rutgers Univ., Ecology, Evolution and Natural Resources, New Brunswick, NJ, USA.*

Ecography

43: 11–24, 2020

doi: 10.1111/ecog.04707

Subject Editor: Carsten Dormann

Editor-in-Chief: Miguel Araújo

Accepted 9 September 2019



www.ecography.org

Species distribution models (SDMs) are a common approach to describing species' space-use and spatially-explicit abundance. With a myriad of model types, methods and parameterization options available, it is challenging to make informed decisions about how to build robust SDMs appropriate for a given purpose. One key component of SDM development is the appropriate parameterization of covariates, such as the inclusion of covariates that reflect underlying processes (e.g. abiotic and biotic covariates) and covariates that act as proxies for unobserved processes (e.g. space and time covariates). It is unclear how different SDMs apportion variance among a suite of covariates, and how parameterization decisions influence model accuracy and performance. To examine trade-offs in covariation parameterization in SDMs, we explore the attribution of spatiotemporal and environmental variation across a suite of SDMs. We first used simulated species distributions with known environmental preferences to compare three types of SDM: a machine learning model (boosted regression tree), a semi-parametric model (generalized additive model) and a spatiotemporal mixed-effects model (vector autoregressive spatiotemporal model, VAST). We then applied the same comparative framework to a case study with three fish species (arrowtooth flounder, pacific cod and walleye pollock) in the eastern Bering Sea, USA. Model type and covariate parameterization both had significant effects on model accuracy and performance. We found that including either spatiotemporal or environmental covariates typically reproduced patterns of species distribution and abundance across the three models tested, but model accuracy and performance was maximized when including both spatiotemporal and environmental covariates in the same model framework. Our results reveal trade-offs in the current generation of SDM tools between accurately estimating species abundance, accurately estimating spatial patterns, and accurately quantifying underlying species–environment relationships. These comparisons

© 2019 The Authors. Ecography published by John Wiley & Sons on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

between model types and parameterization options can help SDM users better understand sources of model bias and estimate error.

Keywords: boosted regression tree, fisheries, generalized additive model, spatial ecology, spatiotemporal model, species distribution model

Introduction

Describing patterns and understanding processes are two goals of spatial ecology. Species distribution models (SDMs) can be used to describe the patterns and processes of species distribution and abundance, and have become essential tools for ecology and conservation (Guisan and Thuiller 2005, Elith and Leathwick 2009). SDMs use abiotic and biotic factors to describe where species occur across a landscape, and in what quantities (Elith and Leathwick 2009). As such, SDMs can provide important insight into critical habitat, connectivity within and across habitats, and the effects of anthropogenic forces on species (Rosenberg et al. 2000, Robinson et al. 2017). SDMs are also increasingly being used to tackle the more difficult problem of forecasting how climate change and other factors may alter range extents, cause distributional shifts, or modify timing of animal migrations (Edwards and Richardson 2004, Hazen et al. 2013, Brown et al. 2016, Thorson et al. 2017). The wide utility of SDMs has given rise to a myriad of SDM types and parameterization options.

SDMs can broadly be categorized into correlative and mechanistic approaches (Robertson et al. 2003, Dormann et al. 2012, Connolly et al. 2017). In correlative models, parameter responses are not pre-defined and are instead modeled implicitly, resulting in responses that are not always ecologically reasonable. Whereas, mechanistic models use explicit functions to characterize relationships among different components of the ecosystem and are usually defined a priori based on ecological theory (Dormann et al. 2012, Connolly et al. 2017). Relative to mechanistic models, correlative models are often conceptually simple, are capable of performing as well or better than simple mechanistic models in estimation or near-term forecasting applications (Robertson et al. 2003, Muhling et al. 2016), and their independence from explicit assumptions can avoid confirmation biases (Connolly et al. 2017). However, correlative models can be limited in their capacity to estimate conditions when system states change or exhibit non-linearity, such as under climate-driven change (Plagányi et al. 2011, 2014, Lurgi et al. 2012). Here we focus on correlative models, as they are presently the most common approach used in SDM applications (Robinson et al. 2017). Indeed, there has been an acceleration in the number of correlative tools available to model species distributions (Elith et al. 2006, Lawler et al. 2006) and there is need to examine the performance, precision and biases of correlative SDMs in a comprehensive manner.

Correlative SDMs use statistical methods to relate species distribution and abundance to abiotic and biotic covariates.

Environmental covariates in correlative SDMs give insight into the underlying processes driving species distributions by indicating potential causal mechanisms. For example, temperature can shape a species distribution by directly acting on its physiological performance (Kearney and Porter 2009), or indirectly by shaping predator distribution in response to prey resources (Harris et al. 2005). If key environmental processes are not accounted for in the model framework because they are unobserved or unmeasured (a common occurrence in ecological studies) then models can be poorly fit. Space and time covariates (e.g. latitude, longitude, day, month, year) can act as proxies for these unmeasured processes, as environmental covariates are usually spatially and temporally autocorrelated (e.g. temperature often exhibits both latitudinal and seasonal gradients) (Legendre 1993). While space and time covariates cannot be used to attribute variation in distribution or density to different top-down or bottom-up mechanisms underlying species' spatial patterns, these covariates can be powerful predictors of species occurrence. Space and time covariates are especially useful where systems are relatively homogenous and slow-changing, where species exhibit strong site fidelity (e.g. central place foragers), or when phenological patterns occur (e.g. annual migrations, spawning behavior) (Bascompte and Solé 1995, Koenig 1999, Dormann 2007).

Parameterization of SDMs with environmental and spatiotemporal covariates should be undertaken with explicit consideration of model purpose (Guillera-Aroita et al. 2015, García-Díaz et al. 2019). For example, if a model's purpose is to accurately quantify species abundance, say for a stock assessment, then quantifying the environmental mechanisms underpinning abundance is less important than accurately estimating abundance. Conversely, if a model's purpose is to predict or project distributions into the medium or distant future, then space and time covariates are less useful in obtaining robust predictions. Although it is intuitive that SDM parametrization should be informed by its intended purpose, it is unclear how different SDM types apportion variance among covariates. Therefore, there is a clear need to assess how SDM performance is influenced by covariate parameterization, and how this varies by the type of model used.

The goal of this paper is to evaluate the performance of three different SDMs and their associated parametrization of spatiotemporal and environmental covariates. To achieve our goal, we use both simulated and empirical data to fit SDMs and evaluate model performance using a suite of seven metrics. Simulated data provides a landscape of species distribution with pre-specified processes shaping species distribution and abundance, whereas processes shaping species distributions are not known a priori for empirical data. Empirical data were obtained from fisheries-independent surveys and thus the case study incorporates additional elements of data uncertainty associated with sampling biases and variation in catchability (Kotwicki and Ono 2019). Comparative analyses of models built on both simulated and empirical data can help identify biases in model performance (e.g. whether a model

systematically underestimates random noise in the data) and challenges in model fitting (e.g. estimation of observation or process error).

We first simulate a species distribution with known environmental preferences, and use this simulated data to fit and compare three common types of SDMs: a machine learning model (boosted regression tree), a semi-parametric model (generalized additive model) and a mixed-effects model (vector autoregressive spatiotemporal model). These three model types represent common correlative modeling approaches for SDM applications, with the latter (VAST) recently becoming an important SDM for fisheries scientists (Thorson 2019b). We then apply the comparative model framework to empirical data for three species (arrowtooth flounder *Atheresthes stomias*, pacific cod *Gadus macrocephalus* and walleye pollock *Gadus chalcogrammus*) in the Eastern Bering Sea, Alaska, USA. These species were chosen to showcase inter-specific variability among results, as all species show individual and complex relationships with the environment and each other (Kotwicky and Lauth 2013, Livingston et al. 2017, Barbeaux and Hollowed 2018). Finally, our discussion focuses on the apparent trade-offs between SDM types and covariate parameterization when estimating species distribution and abundance.

Material and methods

Simulated data

Spatially-explicit abundance of a hypothetical species was simulated over 20 yr in a $20 \times 20^\circ$ gridded area with $1 \times 1^\circ$ grid cells, and used to fit three types of species distribution models (SDMs). We specified that the simulated species distribution is determined by habitat suitability and is not dispersal limited. Data were simulated using two major steps. First, habitat suitability grid layers (1° resolution) were created for each year with four sources of variability: environmental, spatial, temporal and spatiotemporal (Fig. 1), where spatial, temporal and spatiotemporal variability represent proxies for unmeasured ecological processes (see Supplementary material Appendix 1 for more details). Environmental habitat suitability included two commonly used covariates, temperature ($^\circ\text{C}$) and topography (m). Temperature had an additional temporal trend (spatially constant), where temperature gradually increased by 2°C over the 20-yr simulation period, with the species' preference remaining constant. Spatial habitat suitability included two covariates, latitude and longitude. Spatiotemporal habitat suitability had three primary modes that fluctuated between high, medium and low suitability

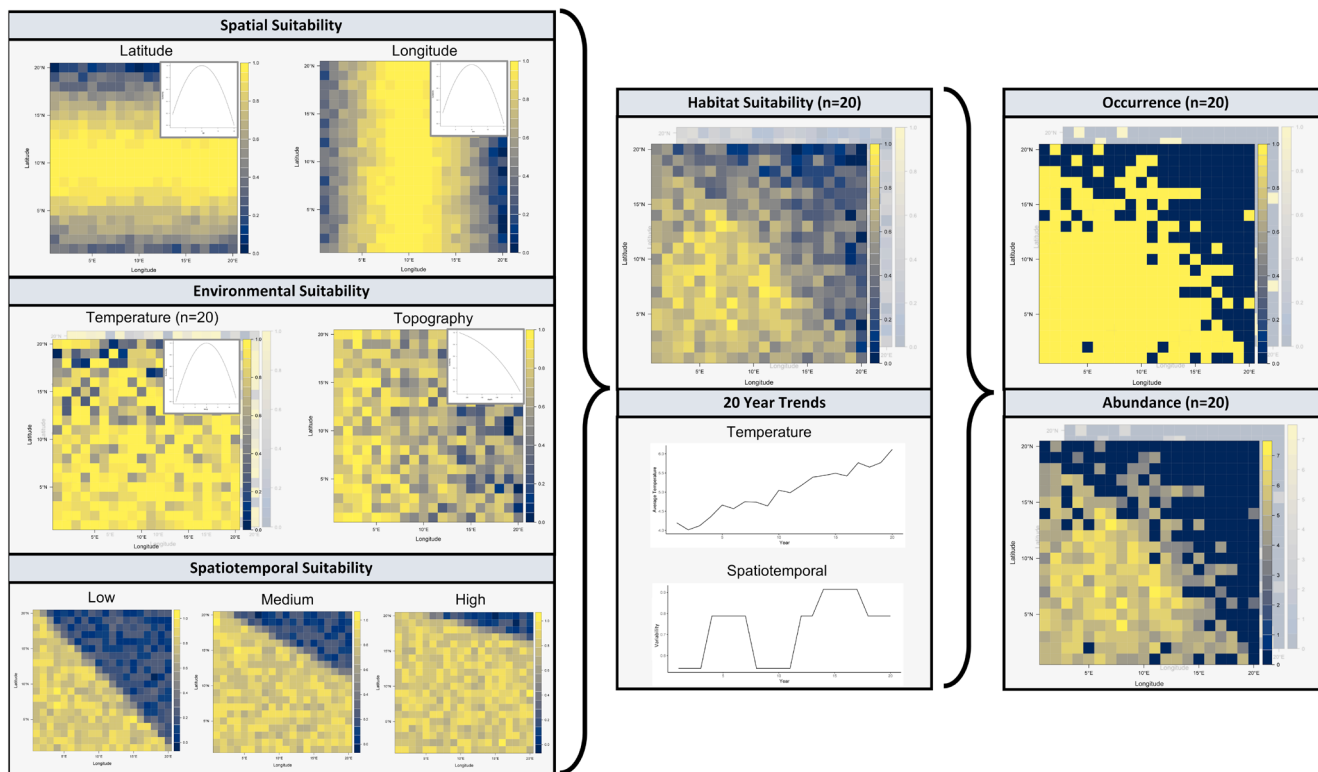


Figure 1. Conceptual outline of how simulated species occurrence and abundance were generated for species distribution models. Simulated data was generated from a lognormal distribution multiplied by a habitat suitability layer. Habitat suitability was the weighted average of spatial suitability (latitude and longitude, single weight), environmental suitability (temperature and topography, double weight) and spatiotemporal suitability (three modes, single weight). Temperature differed each year with an overall increase of 2°C over 20 yr.

across the 20-yr simulation (Fig. 1, Supplementary material Appendix 1). Habitat suitability ranged from 0 to 1, with highest habitat suitability being 1.

Second, these habitat grid layers were used to inform species occurrence and abundance at each grid cell for each year. Habitat suitability for each grid cell was used to determine if a species was present (1) or absent (0) (Supplementary material Appendix 1). If the species was absent then abundance was zero. If the species was present then abundance was estimated from a log-normal distribution with log-mean two and log-standard deviation of 0.1 (Fig. 1, Supplementary material Appendix 1 Table A1). Abundance in each grid cell was then multiplied by the habitat suitability value in the same grid cell, to provide a habitat-informed abundance. Simulated data were generated for each grid cell ($n = 400$) once per year, for 20 yr, resulting in 8000 data points. The simulation used the ‘virtualspecies’ R package (Leroy et al. 2016) that is specifically designed to reflect real-world properties and datasets (Meynard et al. 2019). Detailed methods for the simulation are provided in the Supplementary material Appendix 1, and the code is available on github.

Species distribution model types

We fit three types of SDMs to the simulated species distributions: boosted regression trees (BRTs), generalized additive models (GAMs) and vector-autoregressive spatiotemporal models (VAST). We constructed all models as delta models, where both the probability of occurrence (binomial) and abundance (log-normal) are modeled as individual components. Delta models are a common approach to estimating species distributions, particularly when data is zero-inflated (such as data from surveys, including the fisheries data used in case studies below) or when different processes are expected to influence occurrence and abundance (Martin et al. 2005, Elith et al. 2006). The three model types represent machine learning and statistical models, two common but conceptually different approaches to modeling species distribution and abundance. Machine learning methods (e.g. BRTs) are gaining in popularity in ecology due to their easy implementation, reliable parameter estimation and strong predictive performance (Elith et al. 2008). Machine learning uses an algorithm to learn the relationships between the response and predictor variables. This is in contrast to statistical models (e.g. GAMs and VAST) that first require assumptions about the underlying distribution of the data (e.g. Gaussian data distribution) before estimating parameters based on that data distribution. Statistical models are the most widely used approach in ecological modeling, with an incredibly diverse suite of parameterization options to meet the required statistical assumptions.

The three model types differ in several key ways. First, the models differ in their ability to estimate nonlinear environmental covariate responses. BRTs automatically estimate nonlinear responses through binary splits (regression trees). GAMs can flexibly estimate highly nonlinear responses using smoothing functions. By contrast, VAST currently estimates

linear responses, but can be configured to produce a simple parabolic response by specifying a quadratic term. The VAST model therefore has a less flexible specification for covariate responses compared with either BRTs or GAMs. Second, standard errors for model predictions are generated differently among the three models and are not directly comparable. BRTs do not conventionally produce standard errors but their inherent stochasticity can be leveraged to produce standard errors by fitting multiple models with random subsets of data and computing the error of predictions (Supplementary material Appendix 1); GAM standard errors are a measure of statistical accuracy of model estimates and are generated based on the Bayesian posterior covariance matrix of the model parameters; and VAST standard errors reflect uncertainty about density predictions and are obtained using a generalization of the delta method (Kass and Steffey 1989, Thorson 2019b).

Third, the models differ in their handling of spatial and spatiotemporal autocorrelation that arises when a statistical model does not account for all processes driving species distribution and abundance. BRTs use recursive binary splits to relate the response variable to covariates (Elith et al. 2008, Merow et al. 2014). The spatial GAM component is configured to include a gaussian process smoother on spatial coordinates, where the spatiotemporal component includes this smoother on spatial coordinates as well as its interaction with year. By contrast, VAST estimates a Gaussian Markov random field (GMRF) for spatial variation as a random effect in the spatial model, and a GMRF for spatial variation and its interaction with year in the spatiotemporal model.

Covariate configurations

Three covariate configurations were tested for each of the three model types (Supplementary material Appendix 1 Table A2). The first configuration included spatiotemporal covariates only (latitude, longitude and year), the second included environmental covariates only (temperature, topography and year), and the third included both spatiotemporal and environmental covariates (latitude, longitude, temperature, topography and year). Year was included in all three model configurations due to the simulated data structure, where data was sampled once per year for 20 yr. Spatiotemporal processes were parameterized as the statistical interaction of latitude, longitude and year (Supplementary material Appendix 1 Table A2). Here we consider this spatiotemporal process as a latent factor, where the spatiotemporal process plays a role in species distribution and abundance but the process is not directly observed or included in the model (as is the case in many real-world scenarios). The three-configuration approach allowed the contribution of spatiotemporal and environmental covariates to be quantified for each model type ($n = 3$), as well as to quantify how well models perform when processes are missing in the SDM framework but are replaced by correlated covariates. Model parameterization for each model type is outlined in detail in the Supplementary material Appendix 1, and where possible

we implemented default or commonly used model settings for simplicity (Supplementary material Appendix 1 Table A2).

Model estimates of species abundance were compared to the simulated data, which represents the truth, for each location and each year using two metrics: root mean square error (RMSE) and Spearman correlation coefficient. Model fit was then assessed using two metrics: Akaike's information criterion (AIC), and percent deviance explained. There is no widely accepted approach to calculate AIC for BRTs, and instead we tested how many covariates (if any) could be removed to simplify the model using the 'gbm.simplify' function in R (Elith et al. 2008). Models were also assessed and compared by visually examining model response curves to environmental covariates, spatially-explicit predictions of abundance including calculations of center of gravity and inertia (Wuillez et al. 2009), and time-series of species total abundance. Finally, the generation and modeling of simulated data was replicated ten times. Within each replicate, the root mean square error (RMSE) of model predictions for each location and year relative to the simulated data was calculated for each model. We tested the influence of model type and covariate configuration on model performance (RMSE) using a two-factor ANOVA and a Tukey honestly significant difference (HSD) post-hoc test (Brodie et al. 2018a) in the R 'stats' package (R Core Team).

Case study

The SDM framework that was outlined above was applied to three species in the Eastern Bering Sea: arrowtooth flounder *Atheresthes stomas*, pacific cod *Gadus macrocephalus*, and wall-eye pollock *Gadus chalcogrammus*. These species were chosen to highlight inter-specific variability and add real ecological complexity to the model comparison, such as species-specific ecological processes not captured by model covariates. Species catch data and contemporaneous in situ environmental data were obtained from the National Oceanic and Atmospheric Administration's Alaska Fisheries Science Centre fishery-independent bottom trawl survey, which used a systematic survey design and standardized fishing procedures to sample ~376 stations each year from 1982 to 2017 (see Stauffer 2004, for details about the survey protocol). The survey catch per unit effort (CPUE) is estimated using the area-swept method (Alverson and Pereyra 1969), which multiplies the distance fished by the average distance between wing tips measured using acoustic spread sensors (see Weinberg and Kotwicki 2008, for details). We used the CPUE of each species as a response variable in modeling, where CPUE is the number of individuals 0.01 km⁻² trawl. CPUE data were truncated to only include adult fish (pollock > 40 cm, arrowtooth flounder > 30 cm, pacific cod > 70 cm). Environmental covariates included bottom temperature (°C) and bottom depth (m), where temperature was recorded at the depth of the trawl gear and bottom depth was recorded by the vessel. Model parameterization was the same as the simulation (Supplementary material Appendix 1), except VAST was implemented across 100 locations (instead of 400 used above) that were selected

by applying a k-means clustering algorithm to the CPUE data to identify geographic locations that reflect survey sampling intensity (Johnson et al. 2019). Model fit was assessed using Akaike's information criterion and percent deviance explained. Model outputs were visualized using a time-series of species total abundance.

Data deposition

We used bottom trawl data collected by the Eastern Bering Sea bottom trawl survey, publicly available at: <www.afsc.noaa.gov/RACE/groundfish/survey_data/data.htm>. Code for the simulated data available at: GitHub, <https://github.com/stephbrodie1/Simulation_SDM_TradeOffs>.

Results

Species distribution models using simulated data

Simulated spatially-explicit species abundance data were successfully modeled using three model types (BRT, GAM, VAST) across three model configurations (spatiotemporal, environmental, spatiotemporal and environmental configurations). The ability of the SDMs to replicate the known environmental preferences of the simulated species was best in the BRT and GAM models. GAMs and BRTs very closely replicated the preference curves for temperature (Fig. 2) and topography (Supplementary material Appendix 1 Fig. A2), with both covariates significant in the GAM ($p < 0.001$; Supplementary material Appendix 1 Table A3) and important in the BRT (no covariates dropped during model simplification; Table 1). The VAST delta model was only able to replicate preference curves in either the occurrence or abundance component of the delta model. The VAST occurrence component could produce the correct U-shaped temperature preference but the mean was slightly right skewed (Fig. 2). The VAST abundance component estimated the near-linear topographic preference well (Supplementary material Appendix 1 Fig. A2). Spatiotemporal covariates (latitude, longitude and year) were all significant in the GAM and VAST delta models ($p < 0.05$; Supplementary material Appendix 1 Table A3), and had greater than 8% relative influence in the BRT delta model (Supplementary material Appendix 1 Table A3).

Spatially-explicit predictions of species abundance revealed that all three model types could broadly replicate the spatial structure of the simulated data (Fig. 3). Overall, BRT and GAM predictions slightly over-predicted spatially-explicit species abundance, and VAST slightly under-predicted spatially-explicit species abundance (Fig. 3). Differences between the three model configurations were most obvious in the BRT and GAM delta models, where models with only spatiotemporal covariates generally over-smoothed model predictions (Fig. 3). VAST predictions showed no major differences between configuration one (spatiotemporal covariates) and configuration three (spatiotemporal and environmental covariates), but configuration two (environmental covariates)

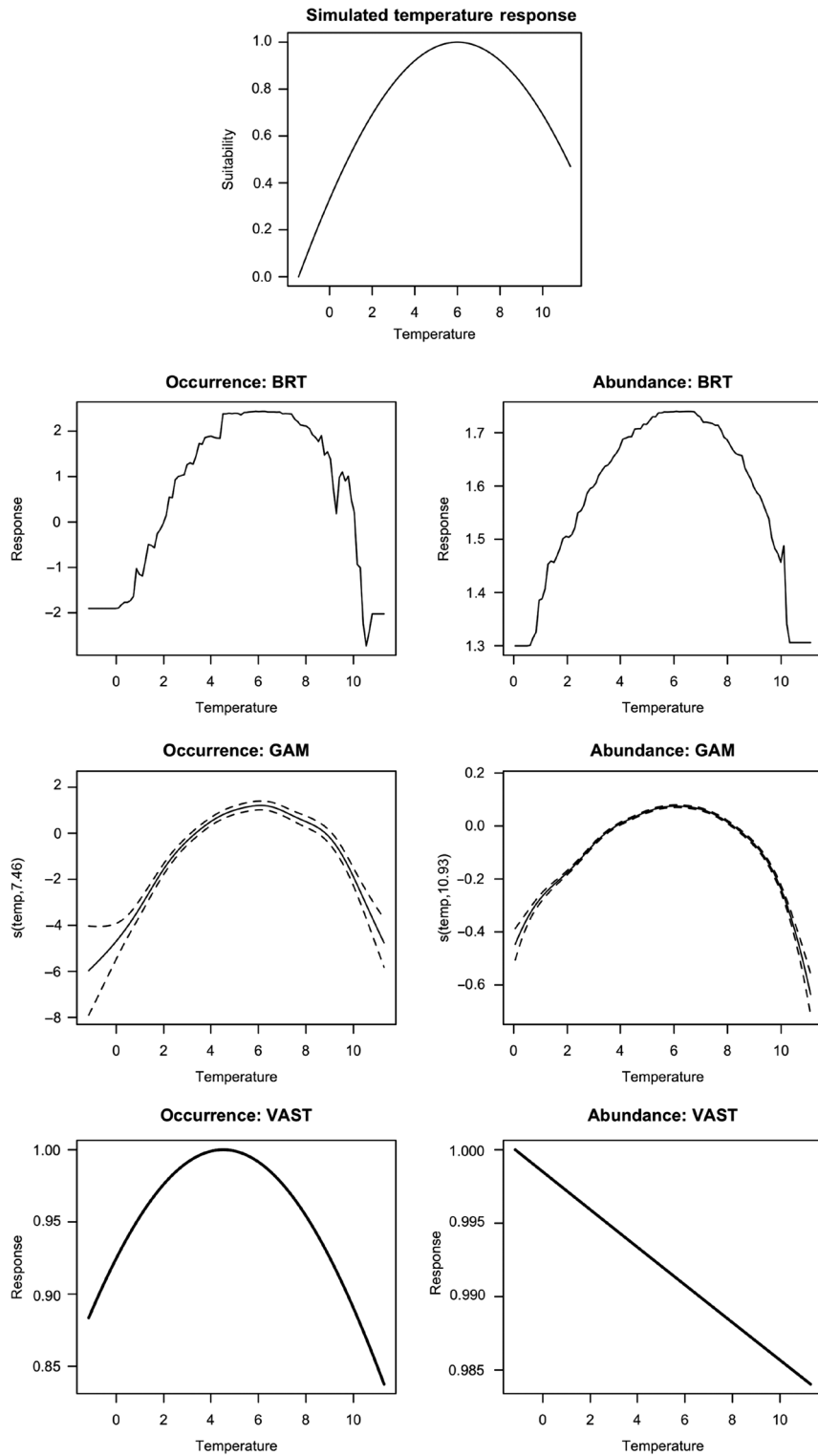


Figure 2. Temperature preference curves for simulated data (top, replicate one) and the three model types (boosted regression tree BRT, generalized additive model GAM and vector autoregressive spatiotemporal model VAST) using configuration three (spatiotemporal and environmental covariates). Models are delta models with occurrence and abundance components presented.

Table 1. Summary statistics for three model-types for simulated (replicate 1) and case study data. Model configurations include spatiotemporal covariates, environmental covariates and the combination of both spatiotemporal and environmental covariates. Root mean square error (RMSE) and spearman correlation coefficient values are shown for the simulated data. Percent explained deviance and Δ Akaike information criteria (AIC) values for each occurrence (abundance) model are shown for simulated and empirical data. AIC values are not generated for each component of the delta model in VAST, instead one value is given. AIC can't be generated for BRTs, and instead we tested how many covariates need to be removed to fit an optimal model (0 indicates no covariates).

Data source	Metric and model	Spatiotemporal	Environmental	Spatiotemporal and environmental
Simulated data	RMSE			
	BRT	1.69	1.26	1.16
	GAM	1.75	1.85	1.23
	VAST	1.35	1.47	1.35
	Correlation			
	BRT	0.79	0.92	0.95
	GAM	0.77	0.77	0.93
	VAST	0.89	0.34	0.89
	Deviance explained			
	BRT	47 (50)	67 (93)	68 (93)
	GAM	43 (41)	36 (50)	62 (86)
	VAST			
	Δ AIC			
BRT	0 (0)	1 (0)	0 (0)	
GAM	1836 (7634)	2479 (6733)	0 (0)	
VAST	9	8168	0	
Arrowtooth flounder	Deviance explained			
	BRT	82 (81)	74 (64)	81 (81)
	GAM	69 (50)	69 (52)	75 (62)
	VAST	NA	NA	NA
	Δ AIC			
	BRT	0 (0)	0 (0)	0 (0)
	GAM	993 (1466)	876 (1249)	0 (0)
VAST	561	1779	0	
Pacific cod	Deviance explained			
	BRT	23 (39)	21 (32)	26 (44)
	GAM	14 (23)	12 (17)	18 (26)
	VAST	NA	NA	NA
	Δ AIC			
	BRT	0 (0)	0 (0)	0 (0)
	GAM	632 (195)	878 (750)	0 (0)
VAST	191	1953	0	
Walleye pollock	Deviance explained			
	BRT	62 (70)	44 (55)	60 (75)
	GAM	30 (42)	32 (38)	41 (48)
	VAST	NA	NA	NA
	Δ AIC			
	BRT	0 (0)	0 (0)	0 (0)
	GAM	902 (1246)	720 (1904)	0 (0)
VAST	320	4776	0	

was not able to reproduce any spatial structure in the simulated data (Fig. 3). Center of gravity and inertia of simulated data were most similar to model predictions that included both spatiotemporal and environmental covariates (Fig. 3).

Total predicted species abundance aggregated by year over the 20-yr time-series indicated major differences between the SDMs. VAST predictions were most precise, and this pattern held for all three model configurations (Fig. 4). However, VAST had the highest standard errors associated with predictions (Fig. 4). BRT predictions were less precise, with delta models over-predicting abundance in some years but errors in predictions were small (Fig. 4). GAM predictions were the

least precise relative to other models, with predictions over- or under-estimating abundance depending on the model configuration (Fig. 4). Standard errors for GAM predictions were often small but overlapped the observed data in some years (Fig. 4). Standard errors of predications are not directly comparable between model types as they are estimated in different ways (see Methods section).

We found that model performance was variable among model types and among covariate configurations (Supplementary material Appendix 1 Fig. A1). The two-way ANOVA indicated that model performance (RMSE) was significantly affected by both model type

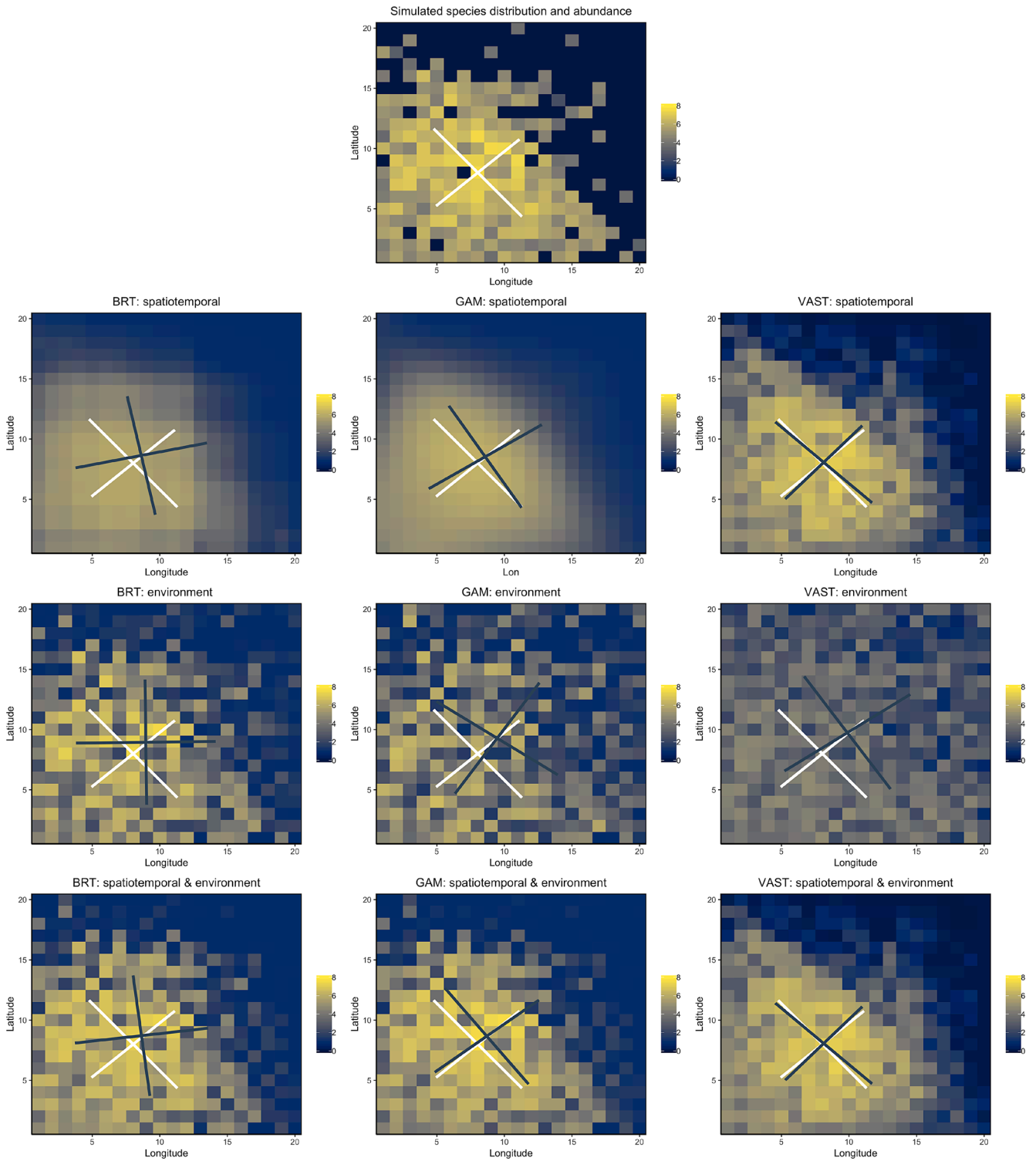


Figure 3. Spatially-explicit species abundance for year one of the simulated data (row 1, replicate one) and delta model predictions with spatiotemporal (row 2), environment (row 3) and both spatiotemporal and environmental covariate configurations (row 4) for each investigated model type. White crosses indicate the true center of gravity and inertia of the simulated data, and grey crosses indicate the predicted center of gravity and inertia from each model.

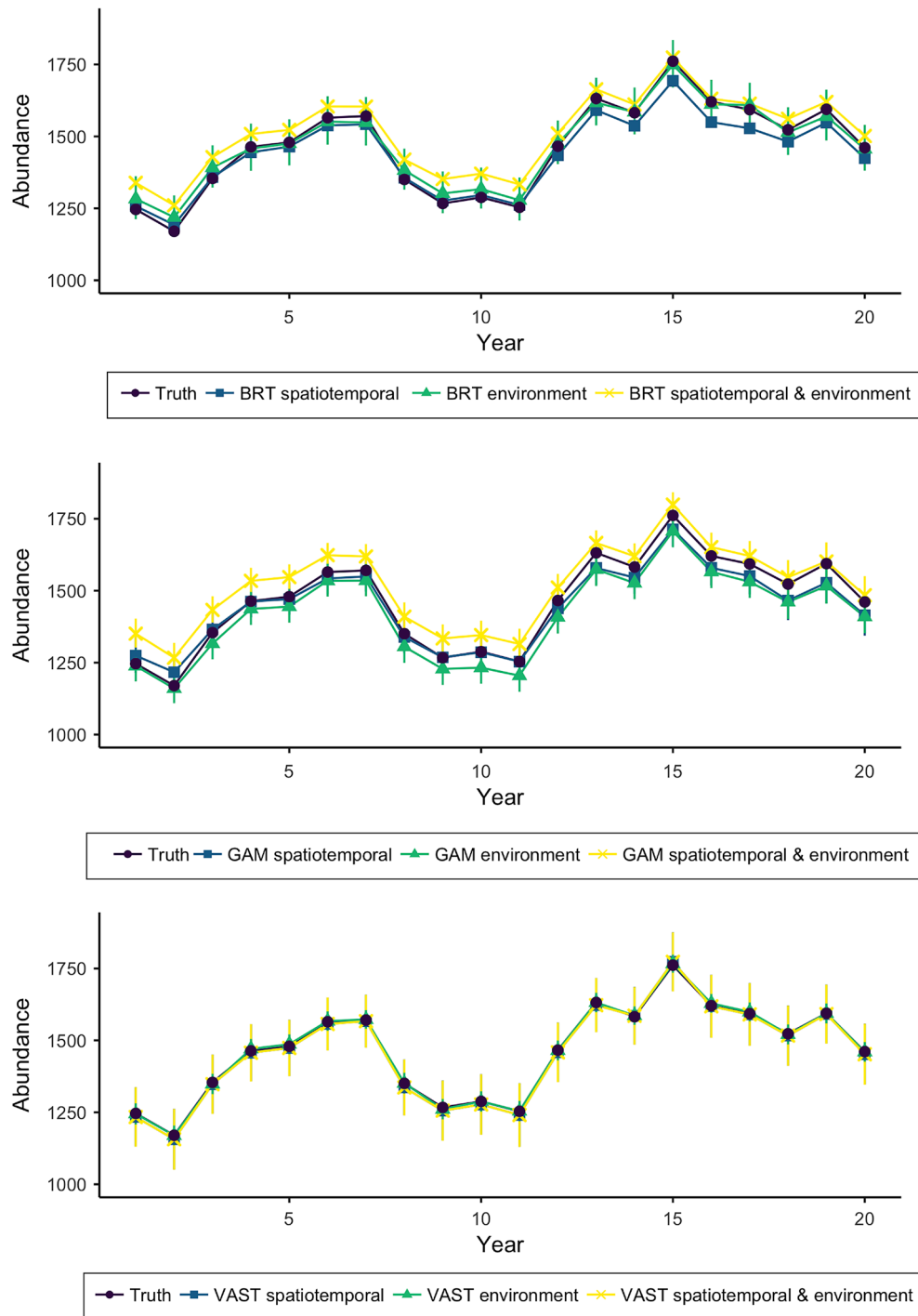


Figure 4. Time-series of simulated (replicate one) and predicted abundance for boosted regression tree (BRT, top), generalized additive model (GAM, middle) and vector autoregressive spatiotemporal model (VAST, bottom) models. Simulated data (purple) and time-series for models with spatiotemporal (blue), environmental (green) and both spatiotemporal and environmental covariate (yellow) are shown. VAST predictions between model configurations are similar, and thus, not all colors are visible. Standard errors are shown for GAM and VAST models, with BRT error determined as the standard deviation of predictions from ten BRTs.

($F_{(2,85)} = 7, p < 0.01$) and covariate configuration ($F_{(2,85)} = 28, p < 0.001$) (Supplementary material Appendix 1 Fig. A1). Post-hoc analyses on the main effect of model type revealed that BRT models were significantly different to GAM

($p < 0.01$) and VAST ($p < 0.01$) models. Post-hoc analyses on the main effect of covariate configuration revealed significant differences between all configurations ($p < 0.01$). Regardless of model type, including both environmental

and spatiotemporal covariates had better explanatory power (explained deviance) and fit (AIC) compared to models that only included either spatiotemporal or environmental covariates (Table 1).

Case study

The distribution and abundance of three species (arrowtooth flounder, pacific cod, walleye pollock) in the Eastern Bering Sea were successfully modeled using all three model types (BRT, GAM, VAST), and the three covariate configurations (spatiotemporal only, environmental only and both spatiotemporal and environmental). A total of 12 866 trawls were included in the analysis for each species, but the percentage of trawls with each species present varied from 43% for arrowtooth flounder, 57% for pacific cod and 90% for walleye pollock (Supplementary material Appendix 1 Table A4). For all three species and three model types, including both spatiotemporal and environmental covariates resulted in the best model fit and highest explanatory power (Table 1, Fig. 5). This result was consistent with that seen for the simulated

data (Table 1). Time-series of species abundance showed BRTs and GAMs predicted lower abundance across all three species compared to VAST estimates and the observed data (Fig. 5). Model estimates for each covariate configuration were typically similar within a model type, except for the arrowtooth flounder GAM, and the VAST models for arrowtooth flounder and walleye pollock (Fig. 5). The similarities between covariate configurations within a model type were consistent with the simulation results.

Discussion

Species are not homogeneously distributed across space and time, and species distribution models (SDM) can be used to better understand the patterns of and processes driving species distributions (Elith and Leathwick 2009, Robinson et al. 2017). Here, we tested a suite of three SDM types (BRTs, GAMs and VAST) and three covariate parameterization options (spatiotemporal covariates only, environmental covariates only and both spatiotemporal and

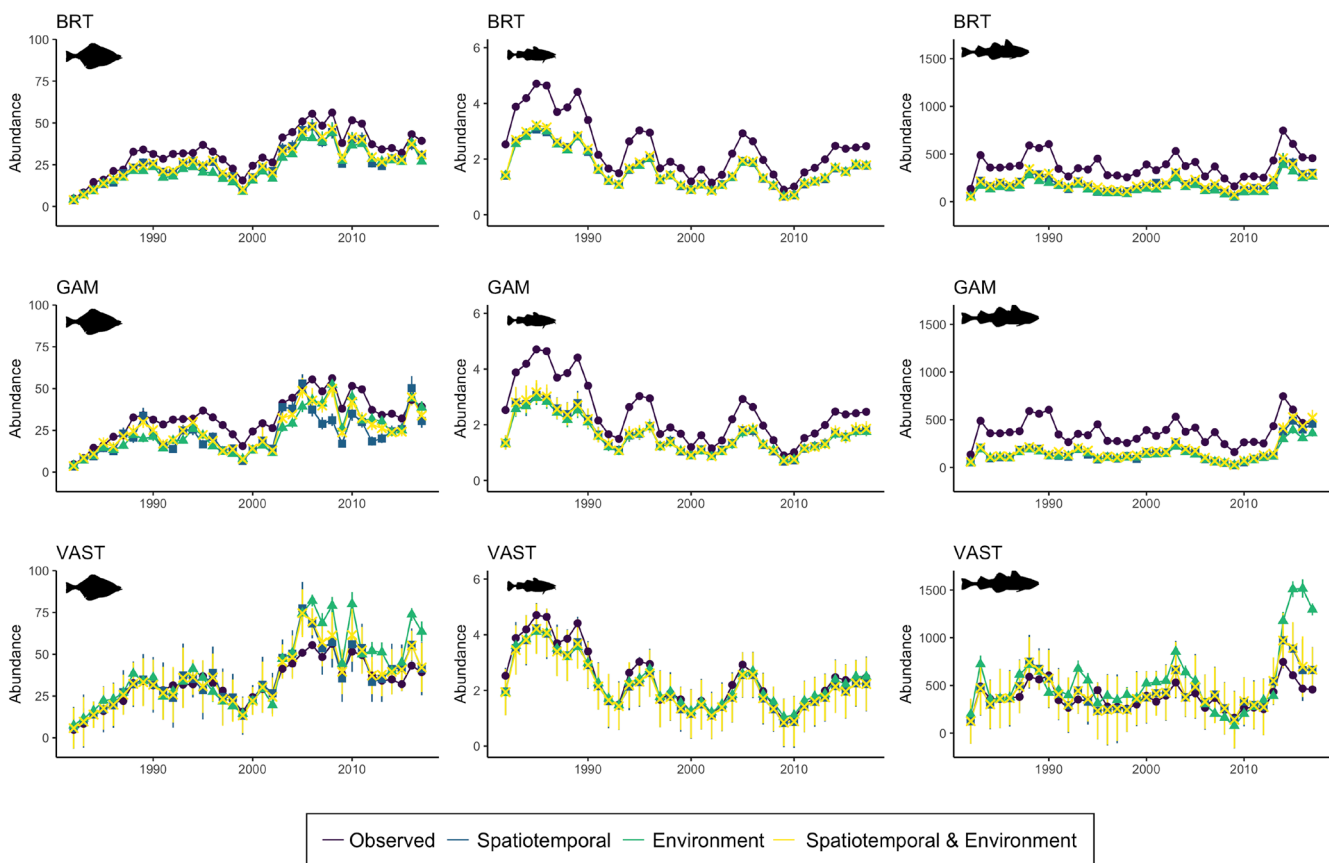


Figure 5. Time-series of predicted abundance of three species in the Eastern Bering Sea arrowtooth flounder (left), pacific cod (middle) and walleye pollock (right). Estimates for boosted regression trees (BRT, top), generalized additive models (GAM, middle), and vector autoregressive spatiotemporal model (VAST, bottom) are shown. Purple is the observed survey data, blue is configuration one (spatiotemporal covariates), green is configuration two (environmental covariates) and yellow is configuration three (spatiotemporal and environmental covariates). Standard errors are shown for GAM and VAST models, with BRT error determined as the standard deviation of predictions from ten BRTs.

environmental covariates), to better understand the influence of model type and covariates on accurately estimating species distribution and abundance. Using simulated and empirical data, we found that model type and covariate parameterization had a large influence on estimates of species distribution and abundance. We found that by including either spatiotemporal or environmental covariates we were typically able to reproduce patterns of species distribution and abundance across the three types of models tested. While including both spatiotemporal and environmental covariates improved model fit and explanatory power across the simulated and empirical SDMs tested, there are trade-offs when including both spatiotemporal and environmental covariates which we discuss below. Our results reiterate the importance of identifying model purpose as a first step in the modeling process. Identifying model purpose can then inform subsequent decisions on which type of model to choose based on the biases associated with each model type, and which configuration of spatiotemporal and environmental covariates to include.

Our simulation indicated that all SDM types could closely replicate the distribution and abundance of a simulated species, however each model revealed varying performance and utility. When including both spatiotemporal and environmental covariates, estimates of simulated spatial distribution were accurate across all three models, but GAMs and BRTs had a smaller error rate than VAST. However, total yearly abundance was most accurately estimated in VAST, followed by BRTs, with GAMs performing worst. Finally, replicating nonlinear environmental preferences was best achieved by GAMs and BRTs, with VAST typically not capable of replicating nonlinear preferences. These results from the simulation have broader implications depending on model purpose. First, if model purpose is to best explain the environmental processes underlying species distributions, then BRTs and GAMs would be a best choice. In such cases, an SDM might be used to predict species distributions in new places or times where only environmental information is available (Hazen et al. 2013, Sequeira et al. 2018, Yates et al. 2018). These kinds of insights are especially important for spatial management and conservation decisions (e.g. related to reducing cross-sector conflicts or placement of protected areas). Second, model purpose may be to obtain robust abundance indices with associated standard errors, in which case VAST would be the best choice. In such cases, the most accurate abundance indices could be included in stock assessments that affect broad-scale decisions on fisheries harvest allowances (Cao et al. 2017, Xu et al. 2019). However, we do note that GAMs and BRTs can still be suitable for deriving abundance indices, with ongoing research indicating that GAMs and machine-learning approaches can have precise predictions with some bias (Stock et al. 2018). Third, user complexity may also affect model purpose. For example, BRTs have the simplest implementation and require minimal statistical assumptions, which are appealing traits when scientific capacity is limited, timelines to decisions are short or both (Tallis et al. 2010). However, such simplicity does

not detract from critical steps of model validation and evaluation in the SDM process (Robinson et al. 2017). Ultimately model purpose must be given due consideration when deciding on SDM type (Guillera-Arroita et al. 2015).

The spatiotemporal and environmental covariate configurations explored in our simulation and case studies touched on a common problem when building SDMs. Including either spatiotemporal or environmental covariates alone can produce good models, but often SDMs are best when both spatiotemporal and environmental covariates are included (Brodie et al. 2018b, Johnson et al. 2019). The simulation and case study shown here support the inclusion of both spatiotemporal and environmental covariates where possible, however this does come with trade-offs using the current generation of SDMs. SDMs used to transfer to another area or another time (i.e. climate projection) may be hindered when spatial and temporal covariates are included. This is a common issue in SDMs where including latitude can greatly improve model explanatory power but limits model transferability to new areas (Becker et al. 2018). A similar problem can occur temporally, when including year as a covariate can limit the capacity of that model to be used in long-term forecasts, such as climate projections. However, the ability of models to transfer beyond the study domain varies between model types as well as model parametrization. For example, gaussian process smoothers will not extrapolate to areas without available data but rather become zero; whereas for other models types extrapolation may need to be explicitly tested and controlled for (Conn et al. 2015, Grüss et al. 2018). Parameterizing models is best done using covariates that have well-established mechanistic underpinnings, especially for models that will be used to transfer to a new area (Yates et al. 2018). However, this will not resolve all issues of unrealistic prediction when extrapolating to unknown conditions and future work could consider the extrapolation methodologies of model types. Additionally, spatiotemporally autoregressive terms within correlative SDMs can provide skill in short-term forecasting (Correia 2018, Thorson 2019c), where forecast skill can come from the persistence of biological and ecological phenomena as well as localized density dependence (Anderson and Beer 2009, Jacox et al. 2017, Mills et al. 2017). As such, model purpose also plays a role in how correlative SDMs should be parameterized.

Identifying model purpose as a first step requires the stakeholders, goals and scope of analysis to be defined prior to model development (Tallis et al. 2010). Deciding on a single model type that is fit for purpose is difficult (Guillera-Arroita et al. 2015), but there are multiple options. These could include building multiple model types and using retrospective model validation and assessment to guide decisions on which model performed best. This requires additional work as many of the models may not ultimately be used but does allow for inter-model uncertainty to be assessed. Alternatively, model ensembles have also been suggested as an effective approach to overcome biases inherent within any one model (Araújo and New 2007, Anderson et al. 2017, Dormann et al. 2018, Abrahms et al. 2019).

Species occurrence and abundance data are common currency for ecologists, but there is large variability among data sources depending on data quality (e.g. opportunistic sampling versus designed surveys), data structure (e.g. daily versus annual or irregular sampling) and data quantity. While not investigated in this paper, such data issues could influence the results shown here. The case study species that we investigated showed similar results across covariate configurations which is likely due to the fact that they were sampled from the same trawl survey, had a high probability of occurrence in each trawl sample (Supplementary material Appendix 1 Table A4), and also share similar life-history traits (Kotwicki and Lauth 2013, Barbeaux and Hollowed 2018). However, the pacific cod models did generally perform worse than the other species, highlighting how variability in species life history may affect model performance across model type. Fitting multiple types of data from different sampling programs (e.g. presence-absence, counts and biomass data) is one way to improve model performance relative to a model that fits only one data source in isolation (Brodie et al. 2018b, Grüss and Thorson 2019). Another option to improve model performance for the case study species would be to add in additional environmental covariates that are known to influence these species distributions (e.g. sediment grain size, cold pool extent, ocean colour, currents, etc.; Parker-Stetter et al. 2016, Laman et al. 2017, Thorson 2019a). This was not done here as our purpose was to provide case study models that were complementary to the simulation, rather than to build the ‘best’ predictive model for an individual species. In general, the total number of covariates included in a model influences whether models are under-fit or over-fit (Merow et al. 2014). The total number of covariates included in a model should depend on the model purpose as well as the understanding of how covariates influence the biological process being modelled (Merow et al. 2014, Fourcade et al. 2018).

Accurately describing and understanding variation in species distribution and abundance is a key requirement for managers and policy makers, often forming the baseline of many strategic plans that seek to conserve resources and improve resilience to future change (NOAA Strategic Plan: NMFS 2016, e.g. Magnuson–Stevens Act: Rosenberg et al. 2000). As such, SDMs form an important part of environmental management and conservation (Robinson et al. 2017). For example, the US National Marine Fisheries Service National Climate Science Strategy explicitly calls for increased focus on understanding relationships between environmental variables and species distributions. This understanding is key to fulfilling mandates for both fisheries sustainability and effective conservation of protected species. With protected species, for example, changes in the regional abundance of a species can influence estimates of potential biological removal within U.S. versus international waters. In fisheries management, accurate estimates of transboundary stock status and productivity interplay within complex social and economic considerations to determine national allocations (Brodziak and Link 2002, Fulton et al. 2011, Bailey et al. 2013). On the US West Coast, specifically, the fisheries management

body (PFMC) is currently moving forward with a ‘climate and communities initiative’ that intends to develop scenarios around shifting distributions of fish stocks and their potential impacts on coastal communities. In another example, the European Union has implemented the Marine Strategy Framework Directive to protect the marine environment and biodiversity by applying an ecosystem approach to management (Brennan et al. 2014). For the above and related efforts, SDMs will form the foundational infrastructure for fulfilling scientific mandates. Our results indicate a trade-off in the current generation of SDM tools between accurately estimating species abundance, accurately estimating spatial patterns and accurately quantifying underlying species–environment relationships. While we found there is no single best model to achieve all these goals, we were able to highlight examples of correlative SDMs that performed well. Additionally, the comparison of model types shown here is of use to resource managers to better understand biases and sources of uncertainty between commonly used models. VAST in particular can very accurately estimate species abundance and associated error (Thorson 2019b), but is a relatively new tool and key differences between VAST and other commonly used tools, like GAMs and BRTs, have not previously been showcased. We recommend further comparisons of VAST and other common tools, as different design constraints will likely affect the relative ranking of VAST and other models (Grüss et al. 2019).

Species distribution models can be used effectively to understand the patterns and processes driving species’ space use and offer robust methods with which to quantify ecosystem components. We investigated covariate parameterization options of correlative SDMs and found these models were capable of accurately representing species distribution and abundances. We identified model purpose as an important step in the SDM process, with decisions on model type and model parameterization shown to have significant effects on model accuracy. The methodological comparison done here can help SDM users better understand sources of model bias and estimate error.

Acknowledgements – This paper is NOAA IEA program contribution #2019_2.

Funding – Contributing support for this work was provided by the National Oceanic and Atmospheric Administration’s Fisheries and the Environment (NOAA FATE) program and NOAA Integrated Ecosystem Assessment (NOAA IEA) Program.

References

- Abrahms, B. et al. 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. – *Divers. Distrib.* 25: doi:10.1111/ddi.12940
- Alverson, D. L. and Pereyra, W. T. 1969. Demersal fish explorations in the northeastern Pacific Ocean – an evaluation of exploratory fishing methods and analytical approaches to stock size and yield forecasts. – *J. Fish. Board Can.* 26: 1985–2001.

- Anderson, J. J. and Beer, W. N. 2009. Oceanic, riverine and genetic influences on spring chinook salmon migration timing. – *Ecol. Appl.* 19: 1989–2003.
- Anderson, S. C. et al. 2017. Improving estimates of population status and trend with superensemble models. – *Fish Fish.* 18: 732–741.
- Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – *Trends Ecol. Evol.* 22: 42–47.
- Bailey, M. et al. 2013. Moving beyond catch in allocation approaches for internationally shared fish stocks. – *Mar. Policy* 40: 124–136.
- Barbeaux, S. J. and Hollowed, A. B. 2018. Ontogeny matters: climate variability and effects on fish distribution in the eastern Bering Sea. – *Fish. Oceanogr.* 27: 1–15.
- Bascompte, J. and Solé, R. V. 1995. Rethinking complexity: modeling spatiotemporal dynamics in ecology. – *Trends Ecol. Evol.* 10: 361–366.
- Becker, E. A. et al. 2018. Predicting cetacean abundance and distribution in a changing climate. – *Divers. Distrib.* 25: 626–643.
- Brennan, J. et al. 2014. EU marine strategy framework directive (MSFD) and marine spatial planning (MSP): which is the more dominant and practicable contributor to maritime policy in the UK? – *Mar. Policy* 43: 359–366.
- Brodie, S. et al. 2018a. Integrating dynamic subsurface habitat metrics into species distribution models. – *Front. Mar. Sci.* 5: 219.
- Brodie, S. et al. 2018b. Citizen science records describe the distribution and migratory behaviour of a piscivorous predator, *Pomatomus saltatrix*. – *ICES J. Mar. Sci.* 75: 1573–1582.
- Brodziak, J. and Link, J. 2002. Ecosystem-based fishery management: what is it and how can we do it? – *Bull. Mar. Sci.* 70: 589–611.
- Brown, C. J. et al. 2016. Ecological and methodological drivers of species' distribution and phenology responses to climate change. – *Global Change Biol.* 22: 1548–1560.
- Cao, J. et al. 2017. Spatiotemporal index standardization improves the stock assessment of northern shrimp in the Gulf of Maine. – *Can. J. Fish. Aquat. Sci.* 74: 1781–1793.
- Conn, P. B. et al. 2015. On extrapolating past the range of observed data when making statistical predictions in ecology. – *PLoS One* 10: e0141416.
- Connolly, S. R. et al. 2017. Process, mechanism and modeling in macroecology. – *Trends Ecol. Evol.* 32: 835–844.
- Correia, H. E. 2018. Spatiotemporally explicit model averaging for forecasting of Alaskan groundfish catch. – *Ecol. Evol.* 8: 12308–12321.
- Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. – *Global Ecol. Biogeogr.* 16: 129–138.
- Dormann, C. F. et al. 2012. Correlation and process in species distribution models: bridging a dichotomy. – *J. Biogeogr.* 39: 2119–2131.
- Dormann, C. F. et al. 2018. Model averaging in ecology: a review of Bayesian, information-theoretic and tactical approaches for predictive inference. – *Ecol. Monogr.* 88: 485–504.
- Edwards, M. and Richardson, A. J. 2004. Impact of climate change on marine pelagic phenology and trophic mismatch. – *Nature* 430: 881.
- Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- Elith, J. et al. 2008. A working guide to boosted regression trees. – *J. Anim. Ecol.* 77: 802–813.
- Fourcade, Y. et al. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. – *Global Ecol. Biogeogr.* 27: 245–256.
- Fulton, E. A. et al. 2011. Human behaviour: the key source of uncertainty in fisheries management. – *Fish Fish.* 12: 2–17.
- García-Díaz, P. et al. 2019. A concise guide to developing and using quantitative models in conservation management. – *Conserv. Sci. Pract.* 1: e11.
- Grüss, A. and Thorson, J. T. 2019. Developing spatio-temporal models using multiple data types for evaluating population trends and habitat usage. – *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsz075>
- Grüss, A. et al. 2018. Producing distribution maps for a spatially-explicit ecosystem model using large monitoring and environmental databases and a combination of interpolation and extrapolation. – *Front. Mar. Sci.* 5: 16.
- Grüss, A. et al. 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. – *Fish. Res.* 213: 75–93.
- Guillera-Aroita, G. et al. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. – *Global Ecol. Biogeogr.* 24: 276–292.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – *Ecol. Lett.* 8: 993–1009.
- Harris, J. E. et al. 2005. Distribution of Gulf of Mexico sturgeon in relation to benthic invertebrate prey resources and environmental parameters in the Suwannee River estuary, Florida. – *Trans. Am. Fish. Soc.* 134: 975–990.
- Hazen, E. L. et al. 2013. Predicted habitat shifts of Pacific top predators in a changing climate. – *Nat. Clim. Change* 3: 234–238.
- Jacox, M. G. et al. 2017. On the skill of seasonal sea surface temperature forecasts in the California Current System and its connection to ENSO variability. – *Clim. Dyn.* 1–15.
- Johnson, K. F. et al. 2019. Investigating the value of including depth during spatiotemporal index standardization. – *Fish. Res.* 216: 126–137.
- Kass, R. E. and Steffey, D. 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). – *J. Am. Stat. Assoc.* 84: 717–726.
- Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – *Ecol. Lett.* 12: 334–350.
- Koenig, W. D. 1999. Spatial autocorrelation of ecological phenomena. – *Trends Ecol. Evol.* 14: 22–26.
- Kotwicki, S. and Lauth, R. R. 2013. Detecting temporal trends and environmentally-driven changes in the spatial distribution of bottom fishes and crabs on the eastern Bering Sea shelf. – *Deep Sea Res. II* 94: 231–243.
- Kotwicki, S. and Ono, K. 2019. The effect of random and density-dependent variation in sampling efficiency on variance of abundance estimates from fishery surveys. – *Fish Fish.* 20: [doi:10.1111/faf.12375](https://doi.org/10.1111/faf.12375)
- Laman, E. A. et al. 2017. Using species distribution models to describe essential fish habitat in Alaska. – *Can. J. Fish. Aquat. Sci.* 75: 1230–1255.
- Lawler, J. J. et al. 2006. Predicting climate-induced range shifts: model differences and model reliability. – *Global Change Biol.* 12: 1568–1584.

- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – *Ecology* 74: 1659–1673.
- Leroy, B. et al. 2016. virtualspecies, an R package to generate virtual species distributions. – *Ecography* 39: 599–607.
- Livingston, P. A. et al. 2017. Quantifying food web interactions in the North Pacific – a data-based approach. – *Environ. Biol. Fishes* 100: 443–470.
- Lurgi, M. et al. 2012. Novel communities from climate change. – *Phil. Trans. R. Soc. B* 367: 2913–2922.
- Martin, T. G. et al. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. – *Ecol. Lett.* 8: 1235–1246.
- Merow, C. et al. 2014. What do we gain from simplicity versus complexity in species distribution models? – *Ecography* 37: 1267–1281.
- Meynard, C. N. et al. 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? – *Ecography*, doi:10.1111/ecog.04385
- Mills, K. E. et al. 2017. Forecasting the seasonal timing of Maine’s lobster fishery. – *Front. Mar. Sci.* 4: 337.
- Muhling, B. A. et al. 2016. Projections of future habitat use by Atlantic bluefin tuna: mechanistic vs. correlative distribution models. – *ICES J. Mar. Sci.* 74: 698–716.
- NMFS 2016. NOAA fisheries protected resources strategic plan: 2016–2020. – <<https://repository.library.noaa.gov/view/noaa/18268>>.
- Parker-Stetter, S. et al. 2016. Factors affecting summer distributions of Bering Sea forage fish species: assessing competing hypotheses. – *Deep Sea Res. II* 134: 255–269.
- Plagányi, É. E. et al. 2011. Modelling climate-change effects on Australian and Pacific aquatic ecosystems: a review of analytical tools and management implications. – *Mar. Freshwater Res.* 62: 1132–1147.
- Plagányi, É. E. et al. 2014. Ecosystem modelling provides clues to understanding ecological tipping points. – *Mar. Ecol. Prog. Ser.* 512: 99–113.
- Robertson, M. P. et al. 2003. Comparing models for predicting species’ potential distributions: a case study using correlative and mechanistic predictive modelling techniques. – *Ecol. Model.* 164: 153–167.
- Robinson, N. M. et al. 2017. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. – *Front. Mar. Sci.* 4: 421.
- Rosenberg, A. et al. 2000. Ecosystem approaches to fishery management through essential fish habitat. – *Bull. Mar. Sci.* 66: 535–542.
- Sequeira, A. M. M. et al. 2018. Transferring biodiversity models for conservation: opportunities and challenges. – *Methods Ecol. Evol.* 9: 1250–1264.
- Stauffer, G. 2004. NOAA protocols for groundfish bottom trawl surveys of the nation’s fishery resources. – NOAA Tech. Memo, NMFS-F/SPO-65.
- Stock, B. C. et al. 2018. The utility of spatial model-based estimators of unobserved bycatch. – *ICES J. Mar. Sci.* 76: 255–267.
- Tallis, H. et al. 2010. The many faces of ecosystem-based management: making the process work today in real places. – *Mar. Policy* 34: 340–348.
- Thorson, J. T. 2019a. Measuring the impact of oceanographic indices on species distribution shifts: the spatially varying effect of cold-pool extent in the eastern Bering Sea. – *Limnol. Oceanogr.* doi:10.1002/lno.11238
- Thorson, J. T. 2019b. Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. – *Fish. Res.* 210: 143–161.
- Thorson, J. T. 2019c. Forecast skill for predicting distribution shifts: a retrospective experiment for marine fishes in the Eastern Bering Sea. – *Fish Fish.* 20: 159–173.
- Thorson, J. T. et al. 2017. The relative influence of temperature and size–structure on fish distribution shifts: a case study on Walleye pollock in the Bering Sea. – *Fish Fish.* 18: 1073–1084.
- Weinberg, K. L. and Kotwicki, S. 2008. Factors influencing net width and sea floor contact of a survey bottom trawl. – *Fish. Res.* 93: 265–279.
- Woillez, M. et al. 2009. Notes on survey-based spatial indicators for monitoring fish populations. – *Aquat. Living Resour.* 22: 155–164.
- Xu, H. et al. 2019. Spatiotemporal dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (*Thunnus albacares*) in the eastern Pacific Ocean. – *Fish. Res.* 213: 121–131.
- Yates, K. L. et al. 2018. Outstanding challenges in the transferability of ecological models. – *Trends Ecol. Evol.* 33: 790–802.

Supplementary material (available online as Appendix ecog-04707 at <www.ecography.org/appendix/ecog-04707>). Appendix 1.