ECOGRAPHY

Research

Climate-informed models benefit hindcasting but present challenges when forecasting species–habitat associations

Cheryl L. Barnes, Timothy E. Essington, Jodi L. Pirtle, Christopher N. Rooper, Edward A. Laman, Kirstin K. Holsman, Kerim Y. Aydin and James T. Thorson

C. L. Barnes (https://orcid.org/0000-0002-2650-5299) ✉ (cheryl.barnes@noaa.gov) and T. E. Essington (https://orcid.org/0000-0002-9986-7100), School of Aquatic and Fishery Sciences, Univ. of Washington, Seattle, WA, USA. – J. L. Pirtle, Alaska Regional Office, National Marine Fisheries Service (NOAA), Juneau, AK, USA. – C. N. Rooper, Fisheries and Oceans Canada, Nanaimo, BC, USA. – E. A. Laman, K. K. Holsman, K. Y. Aydin and J. T. Thorson (https://orcid.org/0000-0001-7415-1010), Alaska Fisheries Science Center, National Marine Fisheries Service (NOAA), Seattle, WA, USA.

Although species distribution models (SDMs) are commonly used to hindcast fine-scale population metrics, there remains a paucity of information about how well these models predict future responses to climate. Many conventional SDMs rely on spatially-explicit but time-invariant conditions to quantify species distributions and densities. We compared these status quo 'static' models with more climate-informed 'dynamic' SDMs to assess whether the addition of time-varying processes would improve hindcast performance and/or forecast skill. Here, we present two groundfish case studies from the Bering Sea – a high latitude system that has recently undergone considerable warming. We relied on conventional statistics (R^2, % deviance explained, UBRE or GCV) to evaluate hindcast performance for presence–absence, numerical abundance and biomass of arrowtooth flounder *Atheresthes stomias* and walleye pollock *Gadus chalcogrammus*. We then used retrospective skill testing to evaluate near-term forecast skill. Retrospective skill testing enables direct comparisons between forecasts and observations through a process of fitting and forecasting nested submodels within a given time series. We found that the inclusion of time-varying covariates improved hindcasts. However, dynamic models either did not improve or decreased forecast skill relative to static SDMs. This is likely a result of rapidly changing temperatures within the ecosystem, which required models to predict species responses to environmental conditions that were outside the range of observed values. Until additional model development allows for fully dynamic predictions, static model forecasts (or persistence forecasts from dynamic models) may serve as reliable placeholders, especially when anomalous conditions are anticipated. Nonetheless, our findings demonstrate support for the use of retrospective skill testing rather than selecting forecast models a priori based on their ability to quantify species–habitat associations in the past.

Keywords: climate change, forecast skill, generalized additive model, hindcast performance, retrospective skill testing, species distribution model

www.ecography.org

Introduction

Marine species have been widely impacted by warming climates. The most notable temperature-driven changes include poleward movements, fluctuations in the amount of suitable habitat and shifts in phenology (Perry et al. 2005, Poloczanska et al. 2013, Thorson et al. 2016, Kleisner et al. 2017). Climate projection models indicate continued ocean warming into the foreseeable future (Wang et al. 2012, Saba et al. 2016, Hermann et al. 2019), compelling re-evaluation of stock units due to persistent shifts in species ranges (Pinsky et al. 2018, Rooper et al. 2020) and realized niches (Silva et al. 2018, Birkmanis et al. 2020, Pennino et al. 2020). When combined, species-specific responses to climate change may give rise to wholesale reorganizations of marine communities via changes in the direction and/or magnitude of species interactions (Jurado-Molina and Livingston 2002, Hunsicker et al. 2011, Albouy et al. 2014, Huntington et al. 2020). Thus, identifying and responding to changes in habitat use is a pressing challenge for fisheries.

Management applications for spatiotemporal analyses are still emerging (Link et al. 2011, Hobday et al. 2016, Berger et al. 2017, Hazen et al. 2018). Scientific interest in using species distribution models (SDMs) to estimate fine-scale population metrics has increased throughout the past century and proliferated in recent years (Guisan and Thuiller 2005, Norberg et al. 2019, Zurell et al. 2020). Specifically, SDMs have become increasingly valuable for identifying environmental drivers of distributions and densities, characterizing relative habitat importance and forecasting shifts in spatial stock structure (Thorson et al. 2017, Laman et al. 2018, Rooper et al. 2020). As the number of SDM applications has grown, so has the need for understanding model performance (Guisan and Zimmermann 2000, Elith et al. 2006, Thorson 2019a, Brodie et al. 2020). Recent work suggests that including both spatiotemporal and environmental covariates maximizes the accuracy of hindcasts (Brodie et al. 2020). Hindcast performance also varies according to the specific modeling framework used, prediction task of interest (e.g. hindcasting versus forecasting) and underlying dynamics of the species in question (Merow et al. 2014, Norberg et al. 2019, Thorson 2019a).

Despite a wealth of information about their descriptive capabilities, we lack a clear understanding of how reliably SDMs forecast species–habitat associations in the near-term. Additionally, many SDMs rely on long-term mean environmental conditions to quantify species distributions and densities (Laman et al. 2018, Pirtle et al. 2019, Kesner-Reyes et al. 2020). Spatially-explicit but time-invariant models, however, do not account for the considerable interannual variability that is prevalent in nature systems; nor can they be used to predict future responses to changing climates. SDMs that lack a temporal component must instead rely on persistence forecasting, which involves carrying predictions forward from the final year used to fit the model. We hypothesized that models with both spatially-explicit and time-varying covariates (referred to as 'dynamic' models henceforth) would outperform those that rely on spatially-explicit but time-invariant covariates (referred to as 'static' models henceforth). Other studies have assessed the implications of hindcasting from different types of SDMs (e.g. maximum entropy, generalized linear models, hierarchical modeling of species communities; Norberg et al. 2019, Brodie et al. 2020). We instead focus on the use of a single modeling framework to compare different treatments of environmental covariates when hindcasting and forecasting fine-scale population metrics. We selected generalized additive models (GAMs) as the basis for our comparison because of their widespread use in fisheries management, including the designation of essential fish habitat in Alaskan waters (EFH; Laman et al. 2018). We also expected nonlinear relationships between each of our focal species and their environments. We refer to hindcasting as predicted species–habitat associations from the past and forecasting as predicted species–habitat associations in years beyond those used to fit models. We relied on well-established statistics to evaluate hindcast performance (i.e. R^2, % deviance explained, UBRE, GCV). We used a less common technique referred to as retrospective skill testing (sensu Thorson 2019a) to assess forecast skill. Retrospective skill testing involves fitting models to some portion of a time series (i.e. a training data set that is treated as the past), predicting population metrics beyond the years used in model fitting (i.e. a testing data set that is treated as the future), and comparing future predictions (i.e. forecasts) to observed values. Directly linking retrospective forecasts to observed population metrics should serve as an improvement over selecting forecast models a priori based on their ability to quantify species–habitat associations in the past.

Material and methods

Overview

We compared four candidate predictor sets when hindcasting and forecasting species–habitat associations (Fig. 1, Supporting information). Our main objective was to assess whether adding time-varying processes to status quo static SDMs would increase hindcast performance and/or forecast skill. Specifically, we evaluated GAMs that accounted for: 1) spatial variation and long-term mean environmental conditions (i.e. static models; S), 2) spatial variation, long-term mean environmental conditions and select time-varying covariates (i.e. simple dynamic models; D1), 3) spatial variation, long-term mean environmental conditions, select time-varying covariates and interannual variability (i.e. intermediate dynamic models; D2) and 4) spatial variation, long-term mean environmental conditions, select time-varying covariates, interannual variability and spatiotemporal variation (i.e. complex dynamic models; D3) (Fig. 1A). We used conventional summary statistics (R^2, % deviance explained, UBRE [unbiased risk estimator] or GCV [generalized cross validation]) to assess the relative performance of hindcast models fit to the entire time series (1982–2018; Fig. 1B). R^2 quantifies the proportion of variance explained by the model, % deviance explained is analogous to unadjusted R^2, and UBRE and
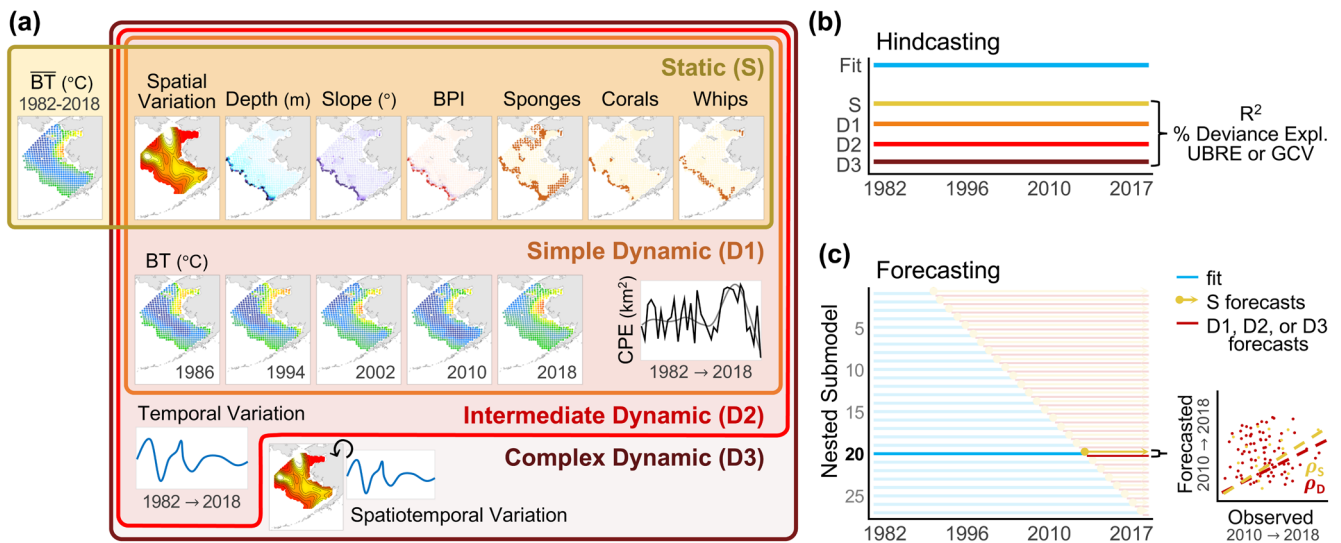
Figure 1. Analytical workflow for hindcasting and forecasting groundfish population metrics in the Bering Sea (1982–2018). (A) Covariate data used to fit species distribution models (SDMs) with varying degrees of complexity (S: static, D1: simple dynamic, D2: intermediate dynamic, D3: complex dynamic). All SDMs accounted for spatially-explicit but time-invariant depth (m), slope (°), BPI (bathymetric position index) and the occurrence of structure-forming invertebrates (i.e. sponges, corals and sea whips). Static SDMs relied on long-term mean bottom temperatures ($\overline{BT}$; °C), whereas all dynamic SDMs accounted for year-specific BT (years displayed represent examples from the full time series). Models D1–D3 included for an interannual index of cold pool extent (CPE; km²). D2 also accounted for temporal variation. D3 incorporated spatiotemporal variation in addition to all other covariates. (B) Time series used for hindcasting. All models (S and D1–D3) were fit (blue) and hindcasted (warm colors) to the entire time series. Performance measures used in model selection are indicated to the right. (C) Time series used to fit (blue) and forecast (S: yellow, D1–D3: red) nested submodels via retrospective skill testing (sensu Thorson 2019a). We estimated Spearman's ρ (i.e. the rank-based correlation between observed and forecasted metrics; example lower right) for each nested submodel and then used a 10-year moving window to estimate near-term (≤ 15 years) forecast skill for each model type (S, D1, D2 and D3).

GCV represent criteria for smoothing parameter estimation in presence–absence and positive catch (i.e. abundance or biomass) GAMs, respectively. We then performed retrospective skill testing (sensu Thorson 2019a) to evaluate forecast skill for each model type. Retrospective skill testing involves fitting a series of models to truncated time series (e.g. 1982–1991 or 1982–2004) and forecasting responses for future years (e.g. 1992–2018 or 2005–2018). We required a minimum of 10 years to fit nested submodels, iteratively added one year to the training data set, and forecasted remaining years in the time series (Fig. 1C). This resulted in 27 nested submodels for each of the two species, three population metrics and four candidate predictor sets (n = 648). We calculated Spearman's correlations (ρ) between observed and forecasted metrics for each nested submodel. We then used a 10-year moving window to quantify mean forecast skill across all nested submodels (Fig. 1C). Analyses were conducted using the statistical programming environment R ver. 4.1 (<www.r-project.org>). Input data and script files can be found at: <https://github.com/cheryl-barnes/SDM-fit-forecast>.

Case study region and species

The Bering Sea is a highly productive marine ecosystem that is characterized by seasonal sea ice and an expansive continental shelf (Stabeno et al. 2001, 2012). We selected the Bering Sea for this work because the region has recently undergone considerable ocean warming and has shown demonstrable shifts in marine species distributions (Stabeno et al. 2017, Stevenson

and Lauth 2019, Huntington et al. 2020, Siddon 2021). The Bering Sea also provides a multidecadal time series of spatially expansive survey data with which to assess changes in fine-scale population metrics (e.g. estimating the distribution of biomass in any given year as opposed to estimating total stock biomass at regional spatial scales). Two economically and ecologically valuable species serve as our case studies: walleye pollock *Gadus chalcogrammus*, a broadly distributed semi-demersal groundfish, and arrowtooth flounder *Atheresthes stomias*, a demersal flatfish that is largely restricted to the middle and outer portions of the continental shelf. Both species are likely to exhibit continued climate-driven shifts in habitat use (Ianelli et al. 2020, Siddon 2021). Although spatiotemporal variation in fishing pressure almost certainly impacts the distributions and densities of harvested stocks (including our case study species), we narrowed our focus to different treatments of environmental covariates to highlight the potential effects of model complexity on environment-only SDMs. We also note that we limited the scope of this study to the use of GAMs. As such, we could not evaluate hindcast performance or forecast skill for other SDMs that are commonly used to quantify species distributions and densities.

Data

Bottom trawl surveys

The Resource Assessment and Conservation Engineering (RACE) Division of the Alaska Fisheries Science Center

(AFSC, National Oceanic and Atmospheric Administration) conducted systematic bottom trawl surveys in the southeastern Bering Sea from 1982 to 2018 (Hoff 2016, Lauth et al. 2019) and in the northeastern Bering Sea in 1982, 1985, 1988, 1991, 2010, 2017 and 2018 (Stevenson and Lauth 2019). AFSC personnel recorded haul-specific numbers and total weights for each species identified. Area swept (km^2) represents survey effort and is calculated as the product of net width (km) and tow distance (km). Net widths were unavailable for some tows in northern Bering Sea. For these, we used a non-linear relationship and previous estimates as starting values (Lauth and Acuna 2007) to predict net width from the length of wire deployed.

We analyzed data from all survey tows to account for presence–absence of arrowtooth flounder and walleye pollock throughout the Bering Sea. We minimized ontogenetic variation in habitat use by constructing GAMs for adult fishes only. We identified adult fish as those measuring greater than the length at 50% maturity (i.e. arrowtooth flounder > 480 mm, Stark 2012, and walleye pollock > 381 mm, Stahl and Kruse 2008). Because fork lengths were measured from a random subsample of 100–300 fish per species and haul, we adjusted haul-specific catch (in numbers or weight) by the proportion of adults subsampled (sensu Barnes et al. 2020). A lack of identification certainty early in the time series required that we combine arrowtooth flounder and kamchatka flounder *Atheresthes evermanni* to represent arrowtooth flounder in the Bering Sea. The closely-related species occupy comparable habitats and have exhibited nearly identical trends in occurrence over time (Wilderbuer et al. 2010). Kamchatka flounder, however, is much less common in the eastern Bering Sea, constituting 7% of the former stock complex (Wilderbuer et al. 2010). We do not expect that the addition of relatively few kamchatka flounder early in the time series would adversely impact model results.

Static habitat covariates
We used a suite of rasters to characterize seafloor habitat throughout the Bering Sea (Supporting information). We interpolated depth (m) from several data compilations (Zimmermann et al. 2013, Zimmermann and Prescott 2018, Mark Zimmermann [AFSC] unpubl., Steve Lewis [AKRO] unpubl., ArcMap, ESRI v10.7) using the natural neighbor method (Sibson 1981). We then used Benthic Terrain Modeler to derive slope (Horn 1981; 3-cell radius) and bathymetric position index (BPI; Guisan et al. 1999; 65-cell radius) from the depth raster (ArcMap; Walbridge et al. 2018) and used these estimates as additional seafloor terrain covariates because of their known effects on fish distributions (Laman et al. 2018, Pirtle et al. 2019). We accessed mean sediment grain size (Φ) from the Eastern Bering Sea Sediment Database (EBSSED; Richwine et al. 2018) and interpolated values using ordinary kriging ('gstat' package in R; Pebesma 2004, 2018) with an exponential model that represented the best fit semi-variogram (Venables and Ripley 2002). In addition to the seafloor metrics previously discussed, Porifera

(sponges), Anthozoa (corals) and Octocorallia (sea pens and sea whips) provide habitat for many groundfish species (Malecha et al. 2005, Stone et al. 2011, Laman et al. 2015). These structure-forming invertebrates (SFIs) also serve as indicators of substrate type (Du Preez and Tunnicliffe 2011), thus we accounted for structure-forming invertebrates as binomial factors in alternative models (Rooper et al. 2014, 2016, Sigler et al. 2015). All static covariate rasters were gridded to a $1-km^2$ resolution and projected to Alaska Albers Equal Area Conic (EAC; standard parallels of 55°N and 65°N; center longitude of 154°W).

Dynamic habitat covariates
Bottom temperature (°C) and cold pool extent (i.e. an isolated body of ≤ 2°C water over the middle shelf; km^2) are well-known ecological drivers in the Bering Sea and are negatively correlated with many groundfish distributions (Rooper et al. 2005, Mueter and Litzow 2008, Stabeno et al. 2012, Thorson 2019b, Grüss et al. 2021). We were interested in comparing models that relied on spatially-explicit but long-term mean bottom temperatures (i.e. the status quo method for Bering Sea SDMs) with those that accounted for spatiotemporal variation in bottom temperature and an interannual index of cold pool extent. We sourced hindcasted bottom temperature and cold pool extent from the BERING10K model that was developed for the northeast Pacific Ocean (Hermann et al. 2013, Kearney et al. 2020). The BERING10K model is a high-resolution, dynamically downscaled, coupled regional ocean modeling system and nutrient phytoplankton zooplankton (ROMSNPZ) model with $10 km^2$ horizontal spatial resolution and 30 depth layers (Hermann et al. 2013, Kearney et al. 2020). BERING10K outputs, which are averaged for each depth and spatial grid cell at a weekly timestep, are available online: <https://beringnpz.github.io/roms-bering-sea>. Bottom temperatures were extracted from the grid cell and week that most closely approximated those of the bottom trawl survey (1982–2018).

We interpolated bottom temperatures using ordinary kriging (gstat package in R; Pebesma 2004, 2018) to similarly link BERING10K hindcasts to a) tow locations from the bottom trawl survey (i.e. model input data) and b) a uniform $1-km^2$ prediction grid spanning the study area. We estimated spatially-explicit but long-term mean temperatures before interpolating for static SDM predictions (Supporting information). We used spatially-explicit and year-specific temperatures to interpolate for dynamic SDM predictions (Supporting information). We then paired interpolated bottom temperatures with the nearest tow location for model fitting and to the nearest point on the uniform grid for predicting and forecasting (sf package in R; Pebesma 2018). We extracted year-specific BERING10K estimates of cold pool extent for fitting dynamic SDMs and modeled the index as a smoothed function of year for predictions. Lastly, we linked static and dynamic covariates to the nearest bottom trawl survey location for model fitting (sf package in R; Pebesma 2018) and to the nearest point on the uniform grid for predicting.

Model fitting

Hindcasting species–habitat associations

We used binomial GAMs to quantify distributions and delta (i.e. hurdle) GAMs to quantify densities of arrowtooth flounder and walleye pollock in the Bering Sea (mgcv package in R; Wood 2011, 2017, Wood et al. 2016). Binomial GAMs rely on presence–absence data to predict probabilities of occurrence. Delta GAMs account for overdispersion in zero-inflated data sets by multiplying probabilities of occurrence by abundance (n) or biomass (kg) predictions from models fit to positive catches. Full static models (S; Eq. 1; Fig. 1A) were specified as follows:

$$f\left(\mu_i\right) = s_1\left(\varphi_i, \lambda_i\right) + s_2\left(z_i\right) + s_3\left(m_i\right) + s_4\left(b_i\right) + s_5\left(\Phi_i\right) \\ + s_6\left(\overline{t}_i\right) + p_i + a_i + o_i + \text{offset}\left(\log k_i\right) \quad (1)$$

where we fit species-specific presence–absence, numerical abundance or biomass μ from haul i using a different distribution and link function f depending on the population metric of interest. We specified presence–absence GAMs with a binomial distribution and complementary log–log link, numerical abundance GAMs with a Poisson distribution and log link, and biomass GAMs with a Gamma distribution and log link. s indicates a tensor product smooth for location (φ: longitude and λ: latitude), represented as a bivariate term to quantify autocorrelated spatial trends. We used low rank isotropic smoothers for depth z, slope m, bathymetric position index b, sediment grain size Φ and long-term mean bottom temperature $\overline{t}$. Additional factors p, a and o represent the occurrence of Porifera (sponges), Anthozoa (corals) and Octocorallia (sea pens and sea whips), respectively. We included log-transformed area swept k as an offset to account for differences in effort among hauls. Overall mean area swept (0.045 ± 0.069 km²) was used as the offset for predicting from static SDMs. All smoothed covariates were estimated using thin plate regression splines and generalized cross-validation (Wood 2003, 2004). We limited substantial changes in predictions across similar values of a particular covariate by restricting the effective degrees of freedom to six for depth and four for all other variables. Effects of location, however, were unconstrained to allow for reasonably complex patterns in habitat use. We also limited the extrapolation of large probabilities outside the range of covariate data by specifying that smoothing penalties should affect the first derivative (i.e. setting m = 1 in 'mgcv'). In doing so, estimated responses would remain constant beyond observed values.

We constructed three dynamic models (Fig. 1A): simple (D1; Eq. 2), intermediate (D2; Eq. 3) and complex (D3; Eq. 4):

$$f\left(\mu_{i,j}\right) = s_1\left(\varphi_i, \lambda_i\right) + s_2\left(z_i\right) + s_3\left(m_i\right) + s_4\left(b_i\right) + s_5\left(\Phi_i\right) \\ + s_6\left(t_{i,j}\right) + g\left(c_j\right) + p_i + a_i + o_i + \text{offset}\left(\log k_i\right) \quad (2)$$

$$f\left(\mu_{i,j}\right) = s_1\left(\varphi_i, \lambda_i\right) + s_2\left(z_i\right) + s_3\left(m_i\right) + s_4\left(b_i\right) + s_5\left(\Phi_i\right) \\ + s_6\left(t_{i,j}\right) + s_7\left(y_j\right) + g\left(c_j\right) + p_i + a_i + o_i \quad (3) \\ + \text{offset}\left(\log k_i\right)$$

$$f\left(\mu_{i,j}\right) = s_1\left(\varphi_i, \lambda_i\right) + s_2\left(z_i\right) + s_3\left(m_i\right) + s_4\left(b_i\right) + s_5\left(\Phi_i\right) \\ + s_6\left(t_{i,j}\right) + s_7\left(y_j\right) + s_8\left(\varphi_i, \lambda_i, y_j\right) + g\left(c_j\right) \quad (4) \\ + p_i + a_i + o_i + \text{offset}\left(\log k_i\right)$$

Simple dynamic SDMs were analogous to static SDMs, except that we replaced long-term mean bottom temperatures $\overline{t}_i$ with year-specific values at each location $t_{i,j}$ and included a spatially varying response g of cold pool extent c. We also replaced overall mean area swept with year-specific means for dynamic SDMs (Supporting information). Intermediate dynamic SDMs extended simple dynamic SDMs by including a smoothed effect of year y (estimated using a cubic regression spline). We added a tensor product interaction of location and year ($\varphi_i, \lambda_i, y_j$) for complex dynamic SDMs, thereby accounting for autocorrelated latent spatiotemporal effects. The entire time series (1982–2018) was used to fit models for hindcasting. We limited concurvity (a generalization of collinearity; mgcv package in R; Wood 2011, 2017, Wood et al. 2016) by excluding covariates with associations ≥ 0.5 and used backward stepwise removal of covariates with a significance threshold of $\alpha = 0.05$ to generate alternative models (Supporting information).

Forecasting species responses to climate change

We used retrospective skill testing (sensu Thorson 2019a) to test whether the addition of time-varying processes would improve forecast skill relative to status quo static SDMs. Retrospective skill testing uses a temporally-blocked design to induce extrapolation (in our case, retrospective forecasts are the extrapolated values). Given that extrapolation is likely to have poor performance relative to interpolation (Roberts et al. 2017), we were interested in assessing the effects of models with varying degrees of complexity on relative hindcast performance and forecast skill. Specifically, we used the best-fit SDMs identified during hindcasting to fit a sequence of nested submodels with variable lengths of the data time series. We required a minimum of 10 survey years to fit each nested submodel and iteratively increased the length of the time series until a single year remained for forecasting (Fig. 1C). In other words, we fit models using data from 1982 to 1991 and forecasted responses from 1992 to 2018. Then we fit models using data from 1982 to 1992 and forecasted responses from 1993 to 2018. We continued fitting and forecasting varied lengths of the time series until the final nested submodel was fit using data from 1982 to 2017 and used to forecast responses in 2018. We applied this approach to static and dynamic SDMs except that static SDMs lacked a temporal component and required the use of persistence forecasts (i.e. forecasts that varied in space but did not change through time).

Assessment and predictions

Hindcasting species–habitat associations

We identified best-fit models as those with the lowest unbiased risk estimator (UBRE; presence–absence models) or generalized cross validation score (GCV; numerical abundance and biomass models). When multiple models generated the same UBRE or GCV, we selected the most parsimonious as best-fit. We then predicted distributions and densities at each point in the uniform prediction grid, once for static SDMs and in each year for dynamic SDMs. We multiplied probabilities of occurrence by numerical abundance or biomass predictions to estimate occurrence-informed densities. We then used model predictions to calculate cumulative population percentiles as a way of characterizing the relative importance of different habitats. We identified areas that encompassed the top 25th percentile of a given population metric as 'hot spots', the top 50th percentile as 'core habitat', the top 75th percentile as 'principal habitat' and the top 95th percentile as EFH (NMFS 2005, Laman et al. 2018).

Forecasting species responses to climate change

We compared model forecasts to observed distributions or densities in each year. We selected Spearman's correlation coefficient (ρ; stats package in R; Hollander and Wolfe 1973, Best and Roberts 1975) as our primary measure of forecast skill because it is a rank-based, nonparametric measure of association that allows for monotonic relationships and relaxes the assumption of bivariate normal distributions (Hauke and Kossowski 2011, Rebekic et al. 2015). Spearman's ρ ignores information about scale and emphasizes a model's ability to identify relative differences in response variables. Spearman's ρ has been used to compare SDMs for data-rich species in other regions (Ready et al. 2010) and can serve as a suitable basis for estimating the degree of confidence in forecasts.

After estimating Spearman's ρ from each nested submodel, we used a 10-year moving window to calculate mean correlations for each species, population metric and model type. Doing so allowed us to assess near-term forecast skill and associated estimates of uncertainty regardless of the length of the time series used to fit models. A 10-year moving window meant that skill for 5-year forecasts represented the mean correlation from 1- to 10-year forecasts. Similarly, skill for 10-year forecasts represented the mean correlation from 6- to 15-year forecasts and skill for 15-year forecasts represented the mean correlation from 11- to 20-year forecasts.

Exploring potential effects on forecast skill

We explored potential reasons for model-based differences in forecast skill by comparing Spearman's ρ from simple, intermediate and complex dynamic SDMs with modified versions of the same models. Modified dynamic SDMs matched original specifications except for the treatment of temperature, where we replaced time-varying bottom temperatures with long-term means. We also estimated Spearman's ρ using persistence forecasts from original dynamic SDMs. This allowed us to compare forecast skill from models that were fit to dynamic covariates but forecasted in a static manner (i.e. assuming a constant response following the final year used in model fitting).

Apart from model-based differences in forecast skill, we were also interested in evaluating spatiotemporal patterns in estimation bias. Thus, we calculated standardized residuals r_s such that: $r_s = r/sd_r$, where $r = x_o - x_f$, x_o represents observed population metrics (presence–absence, numerical abundance or biomass) and x_f represents forecasted population metrics in the same year and location. For simplicity, we only evaluated residuals from-1 year forecasts (i.e. forecasts representing one year into the future).

Results

Hindcasting species–habitat associations

Model performance was directly proportional to model complexity (Table 1, Supporting information). This was true for both species and all population metrics, where R^2 and % deviance explained increased from static to complex dynamic models (S < D1 < D2 < D3) and UBRE or GCV decreased from static to complex dynamic models (S > D1 > D2 > D3). Spearman's ρ also increased with model complexity, suggesting that dynamic SDMs better distinguished historical differences in habitat use than status quo static SDMs (Table 1). Thus, we identified complex dynamic models (D3) as the best-fit SDMs for hindcasting distributions and densities of our case study species. Generally, models (regardless of type or population metric) were more effective at describing variation in habitat use by arrowtooth flounder than walleye pollock (Table 1). Models also explained a greater proportion of variance in distributions than densities, though density-based correlations were slightly higher (Table 1).

We found that arrowtooth flounder are largely restricted to the middle and outer shelf of the Bering Sea, though their distributions expanded toward the latter portion of the time series (Fig. 2, Supporting information). Walleye pollock were more broadly distributed throughout the region and exhibited considerable spatiotemporal variation in relative habitat importance. This was most evident from interannual shifts in the location of hot spots and areas identified as EFH (Fig. 2, Supporting information). Walleye pollock were found in relatively shallow habitats (0–300 m) with a wide range of bottom temperatures (−2.1 to 12.8°C), whereas arrowtooth flounder inhabited deeper depths (between 200 and 600 m) that were always cooler than 6.1°C. The distributions and densities of both species generally decreased with increasing cold pool extent. All partial covariate effects and year-specific population percentiles can be found in the Supporting information.

Forecasting species responses to climate change

Although complex dynamic models outperformed all others when hindcasting, they failed to increase and even decreased

Table 1. Performance measures used when hindcasting arrowtooth flounder *Atheresthes stomias* and walleye pollock *Gadus chalcogrammus* population metrics in the Bering Sea (1982–2018). Types of species distribution models (SDMs): static (S), simple dynamic (D1), intermediate dynamic (D2) and complex dynamic (D3). Covariates included in each model type: S – spatial variation, long-term mean bottom temperature and static habitat covariates; D1 – spatial variation, location- and year-specific bottom temperature, interannual index of cold pool extent and static habitat covariates; D2 – covariates included in D1, plus temporal variation; D3 – covariates included in D2, plus spatiotemporal variation.

	Arrowtooth flounder				Walleye pollock			
	S	D1	D2	D3	S	D1	D2	D3
(a) Probability of occurrence								
R^2	0.612	0.649	0.721	**0.736**	0.342	0.428	0.459	**0.496**
% Deviance Exp.	55.332	59.916	67.182	**69.454**	32.563	41.155	44.530	**48.499**
UBRE or GCV	−0.396	−0.454	−0.551	**−0.573**	−0.456	−0.521	−0.547	**−0.571**
Spearman's ρ	0.751	0.772	0.801	**0.806**	0.466	0.505	0.519	**0.532**
(b) Numerical abundance								
R^2	0.154	0.166	0.193	**0.224**	0.095	0.104	0.117	**0.147**
% Deviance Exp.	33.991	36.277	41.288	**46.464**	24.672	27.573	29.766	**34.975**
UBRE or GCV	22.058	21.270	19.523	**17.737**	944.201	907.806	880.290	**814.931**
Spearman's ρ	0.769	0.789	0.815	**0.822**	0.612	0.668	0.690	**0.711**
(c) Biomass								
R^2	0.148	0.152	0.160	**0.185**	0.077	0.084	0.054	**0.093**
% Deviance Exp.	37.764	39.932	46.955	**51.453**	24.419	28.299	31.163	**35.945**
UBRE or GCV	0.934	0.911	0.807	**0.754**	2.113	2.012	1.935	**1.816**
Spearman's ρ	0.772	0.790	0.821	**0.827**	0.578	0.635	0.658	**0.689**

forecast skill for the species and population metrics we tested (Fig. 3, Supporting information). We found that forecast skill was comparable among the various SDMs at the shortest time step (i.e. five years into the future). Dynamic SDMs, however, displayed a sharper decline in forecast skill as population metrics were predicted further into the future. Like hindcast performance, we found greater forecast skill for arrowtooth flounder compared to walleye pollock. Among the population metrics tested, biomass forecasts were most correlated with observations and showed the least amount of model-based degradation in skill through time (Fig. 3).

Exploring potential impacts on forecast skill

We found that considerable differences in the bottom temperatures used to fit and forecast species–habitat associations may have hindered forecast skill by requiring models to extrapolate temperature-driven responses beyond observed covariate ranges (Supporting information). Specifically, time-varying temperatures used to forecast from dynamic SDMs spanned a much wider range of values than long-term means used to forecast from static SDMs (Fig. 4). Temperature disparities were exaggerated for models that quantified positive catch data (i.e. numerical abundance and biomass), with arrowtooth flounder showing the greatest divergence due to a narrower thermal experience throughout the time series. Long-term mean temperatures used to forecast from static SDMs dampened interannual variability and were more likely to fall within the range of those observed (Fig. 4). When we substituted time-varying temperatures with long-term means (i.e. D1 added only cold pool extent to static models, D2 added a year effect to D1 and D3 added a spatiotemporal term to D2), forecast skill for dynamic SDMs increased to that of static SDMs (Supporting information). We observed the same effect when evaluating persistence forecasts from dynamic SDMs. The exceptions to

this were arrowtooth flounder abundance (models D1, D2 and D3), walleye pollock abundance (model D3 only) and walleye pollock biomass (model D3 only).

Spatiotemporal patterns in standardized residuals showed that prediction bias was similar among static and complex dynamic models (Fig. 5). Over-predictions were more prevalent than under-predictions and differences in magnitude were more prevalent than differences in direction. Biases were largely collocated such that the specific locations of over- and under-prediction were similar among SDMs (walleye pollock in 2003, arrowtooth flounder in 2010, both species in 2017; Fig. 5, Supporting information). There were some years, however, with considerable model-based differences in the degree of bias. For instance, forecasts from static SDMs showed greater over-predictions of walleye pollock biomass in 2010 compared to dynamic SDMs in the same year. Conversely, forecasts from dynamic SDMs showed greater under-predictions of arrowtooth flounder biomass in 2003 compared to static SDMs in the same year.

Discussion

We compared status quo static SDMs with more climate-informed dynamic SDMs to assess whether the addition of time-varying processes would improve hindcast performance and/or forecast skill of fine-scale population metrics. Because SDMs that account for both spatially-explicit and time-varying processes should more closely reflect the degree of variation in nature, we expected forecasts from dynamic SDMs to display greater skill compared to those that rely on long-term mean environmental conditions. Using two groundfish case studies from the Bering Sea, we found that complex dynamic models (i.e. GAMs that accounted for spatial, temporal and spatiotemporal variation) outperformed all others when
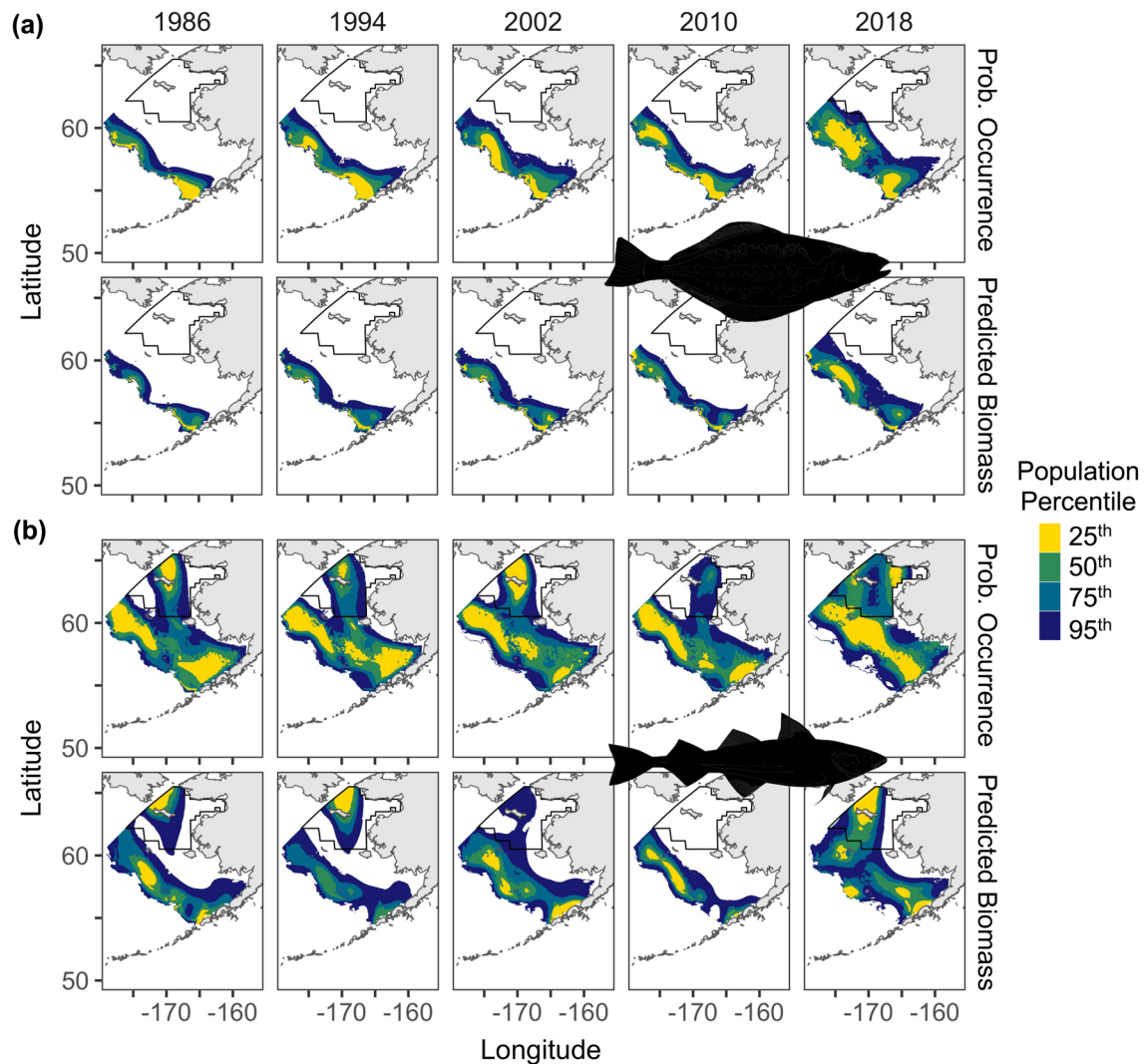
Figure 2. Relative habitat importance based on probability of occurrence and biomass predictions from best-fit species distribution models (SDMs) for arrowtooth flounder (top) and walleye pollock (bottom) in the Bering Sea (select years between 1982 and 2018; see the Supporting information for the full time series). Climate-informed (i.e. 'complex dynamic') SDMs accounted for spatial, temporal and spatiotemporal variation in addition to location- and year-specific bottom temperature, an interannual index of cold pool extent (km²; modeled as a spatially-varying coefficient), and spatially-explicit habitat covariates (depth, slope, bathymetric position index and occurrence of structure-forming invertebrates). Colors illustrate population percentiles: yellow is the top 25th percentile (categorized as 'hot spots'), green is the top 50th percentile ('core' habitat), teal is the top 75th percentile ('principal' habitat) and blue is the top 95th percentile (essential fish habitat [EFH]). The black outlined polygon denotes the northern Bering Sea, which was inconsistently sampled throughout the time series.

hindcasting. However, these more climate-informed SDMs did not improve and, in some cases, decreased forecast skill in the near-term (e.g. 10–15 years into the future) relative to persistence forecasts from status quo static SDMs.

The decline in forecast skill for dynamic SDMs was likely due to the prevalence of forecast temperatures that were outside the range of those used in model fitting. Specifically, insufficient variation in the bottom temperatures used to fit positive catch models (especially for arrowtooth flounder) may have led to inaccurate response curves. SDMs for walleye pollock, for instance, were fit to a much wider range of temperatures than SDMs for arrowtooth flounder. SDMs for walleye pollock also showed less degradation in forecast skill

under novel conditions. These species-specific disparities in temperature were likely exacerbated when models were fit to data from one ecosystem state (e.g. characterized by anomalously cold temperatures) and forecasted in another (e.g. characterized by anomalously warm temperatures). The Bering Sea oscillated between warm and cool phases throughout our time series (Siddon 2021), causing a mismatch between ecosystem states for some of the nested submodels that we ran during retrospective skill testing. Nonetheless, shifts in ecosystem characteristics represent a probable scenario for those interested in forecasting, especially given the increased frequency and intensity of anomalous conditions such as those generated during marine heat waves (Pershing et al. 2019).
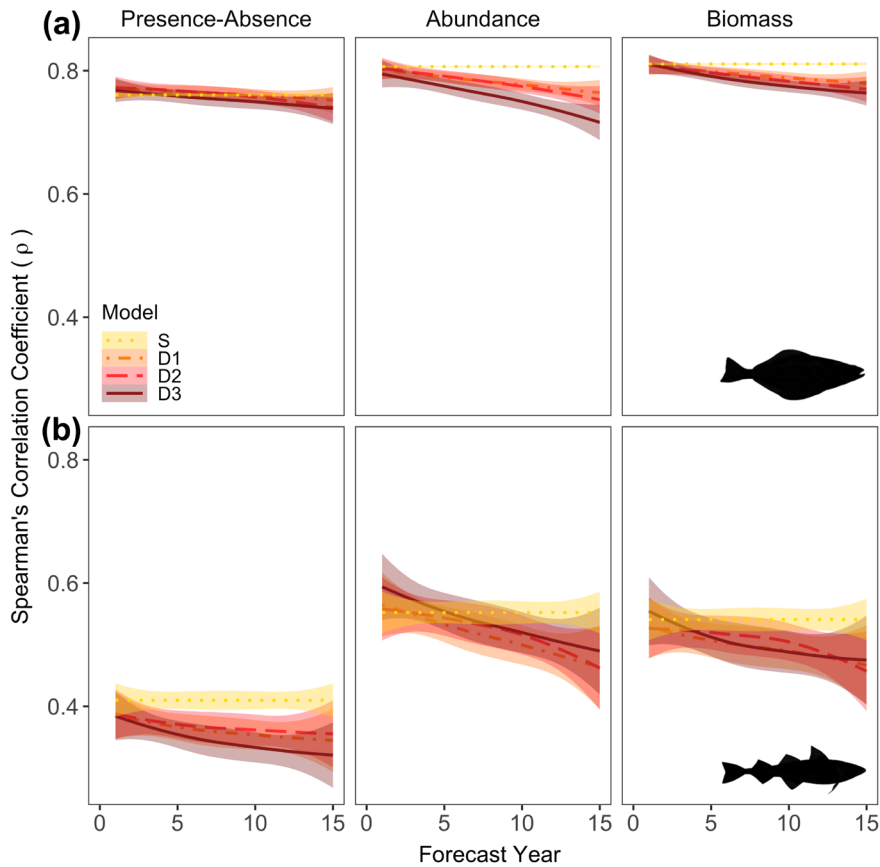
Figure 3. Spearman's correlation coefficients (ρ) used to assess forecast skill from species distribution models (SDMs) for arrowtooth flounder (top) and walleye pollock (bottom) in the Bering Sea (1982–2018). Means (lines) and standard deviations (bands) were calculated using a 10-year moving window for each population metric (presence–absence [left], numerical abundance [center] and biomass [right]) and model type (S: static [dotted, gold], D1: simple dynamic [dot-dash, orange], D2: intermediate dynamic [dashed, light red], D3: complex dynamic [solid, dark red]). Covariates included in each model type: S – spatial variation, long-term mean bottom temperature and static habitat covariates; D1 – spatial variation, location- and year-specific bottom temperature, interannual index of cold pool extent and static habitat covariates; D2 – covariates included in D1, plus temporal variation; D3 – covariates included in D2, plus spatiotemporal variation.

Recommendations for SDM users

We emphasize the need for analysts to consider desired outcomes prior to model specification and to fit for the intended purpose. Our results suggest that SDMs with greater complexity are most appropriate when scientific or management interests are primarily descriptive (i.e. the greater need is to hindcast species–habitat associations). If research objectives involve forecasting, however, model selection should involve some sort of validation process to evaluate predictive skill.

We were surprised that complex dynamic SDMs, which outperformed all others when hindcasting, did not improve and even decreased forecast skill relative to static SDMs that lacked a temporal component. Relatively greater forecast skill from static models contrasts previous work that found increased skill when accounting for dynamic processes (Thorson 2019a). These apparent contradictions may result from different modeling frameworks (i.e. we relied on GAMs whereas Thorson 2019a used a vector autoregressive spatiotemporal [VAST] model). They may also indicate that

forecast skill is scale dependent. We quantified fine-scale distributions and densities whereas Thorson (2019a) assessed population center-of-gravity, a metric that averages across varying degrees of forecast skill at finer spatial scales. Scale-dependent forecasting is consistent with recent conclusions that localized species responses are more nuanced than some of the commonly used (but much coarser) population metrics suggest (Barnett et al. 2021).

To address the context-specific nature of forecasting, we advocate for the routine application of retrospective skill testing. Using this type of validation process allows SDM users to maximize forecast skill and customize model selection for individual species, systems and population metrics. Importantly, our results illustrate how analysts should not assume that past relationships will serve as suitable predictors for the future. Although we focused on GAMs and did not test other SDMs as part of this work, any model-specific discrepancy between hindcast performance and forecast skill signals a need for analysts to test whether forecast models can be identified based on hindcast performance alone. Comparing relative hindcast performance and forecast skill for various
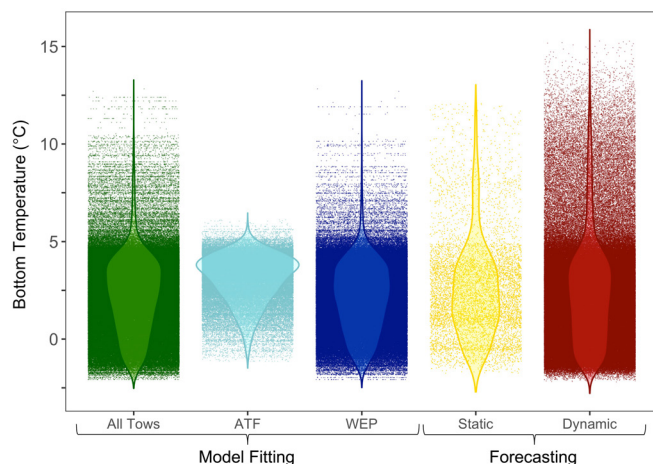
Figure 4. Bottom temperatures (°C) used to fit or forecast ground-fish population metrics from species distribution models (SDMs) in the Bering Sea (1982–2018). Data used to fit presence–absence models (green) reflect year-specific temperatures nearest to bottom trawl survey tow locations (Alaska Fisheries Science Center, NOAA; Lauth et al. 2019). Data used to fit numerical abundance and biomass models were limited to locations with positive catches of arrowtooth flounder (ATF; light blue) and walleye pollock (WEP; dark blue) in any given year. Forecasts from static SDMs (yellow) relied on spatially-explicit but long-term mean temperatures, whereas forecasts from dynamic SDMs (red) were based on spatially-explicit and time-varying bottom temperatures. Shapes illustrate mirrored densities for each group. Data were derived from regional ocean modeling system (ROMS) hindcasts (NOAA's Alaska Climate Integrated Modeling Project; Hermann et al. 2013, Kearney et al. 2020). Bottom temperatures used in all nested sub-models during retrospective skill testing are shown (model fitting from 1982 to 1991, 1982 to 1992, … and 1982 to 2017; forecasting from 1992 to 2018, 1993 to 2018, … and in 2018).

types of SDMs will improve our understanding about how generalizable the need for retrospective skill testing or a similar model selection procedure may be. One such extension of our study would be to evaluate the relative effects of model selection (i.e. selecting forecast models a priori based on hindcast performance or posteriori via retrospective skill testing) for parametric, non-parametric and Bayesian modeling frameworks. In particular, joint SDMs may yield different results given that simultaneously modeling two or more species tends to improve hindcast performance relative to single species SDMs (Ovaskainen and Soininen 2011, Ovaskainen et al. 2017). In addition to retrospective skill testing, we suggest that SDMs users (specifically those who rely on GAMs) consider the use of persistence forecasts for dynamic SDMs as a way of balancing the greater ecological realism that comes from modeling spatiotemporal processes with the more reasonable forecasts generated by static SDMs.

Given current limitations, we caution against the use of absolute measures of forecast skill. Simulation studies that focus on the strengths and weaknesses of different measures of skill could, however, be used to develop absolute estimates that would increase the utility of forecasts for

fisheries management (Conn et al. 2015, Norberg et al. 2019, Allyn et al. 2020). Without these analytical developments, we must continue to rely on tracking relative changes in habitat use (e.g. using population percentiles to characterize EFH, principal or core habitat and hot spots) to reduce the impact of over- or under-predictions. We also note that static SDMs seem to offer realistic near-term forecasts for fine-scale distributions and densities in the Bering Sea, which represents a high-latitude system that has experienced rapid climate change. The relative skill of static model forecasts may have resulted from unidentified shortcomings of our dynamic SDMs (e.g. not accounting for fishing pressure or species interactions as additional time-varying covariates). We also acknowledge the possibility that local scale responses may not have changed directionally and that species–habitat associations were fairly stable within the timeframe we evaluated, making it difficult to improve forecasts with time-varying covariates.

Avenues for continued SDM development

Although we found that modeling dynamic processes failed to improve forecast skill relative to status quo static SDMs, continuing to explore more climate-informed dynamic SDMs and evaluate skill for more dynamic forecasts will help move the field forward. Harris et al. (2018) described forecasting in ecology as being in its infancy. Following weather forecasting as an example, the authors assert that we must continue to assess and refine a variety of approaches to make progress in this realm. Near-term forecasting will be especially useful for improving our ability to predicting species responses to changing environments because 1 to 5 year forecasts can be routinely validated and models expediently fine-tuned.

One research direction that may help us improve forecasts from dynamic SDMs would be to address effects of data quality on forecast skill. For instance, we found that poor forecast skill was largely due to positive estimation bias (e.g. over-predicting biomass relative to observed values), which is of particular concern because overpredicting stock abundance can lead to management advice that does not sufficiently protect against overfishing. The locations of over-predictions appeared to coincide with decreased skill in the BERING10K hindcasts of bottom temperature (Kearney 2021). A formal investigation into the relationship between BERING10K hindcast skill and SDM prediction bias would elucidate how the quality of covariate data may have impacted forecast skill in the Bering Sea. Such an analysis would also guide spatial or temporal limits to extrapolation; for example, by highlighting years and/or locations where covariate skill is below some desired threshold. In the Bering Sea, this could mean eliminating portions of the outer shelf, shelf break and waters surrounding Saint Matthew Island, where BERING10K hindcasts of temperature exhibit poor skill (Kearney 2021). It may also be necessary to account for variables that relate to but are not explicitly accounted for by bottom temperature and cold pool extent (e.g. zooplankton abundance, Hermann et al. 2019; relative foraging rates, Holsman et al.
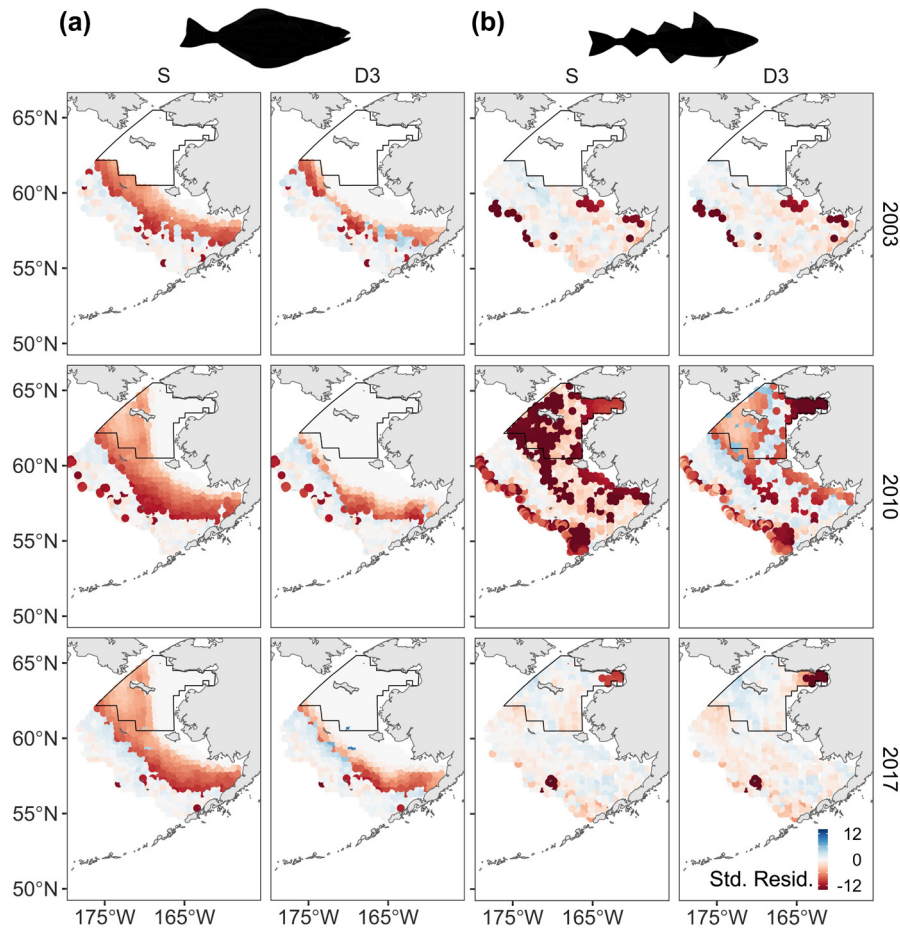
Figure 5. Standardized residuals ($r_s = r/sd_r$, where $r = x_o - x_f$, x_o is observed biomass and x_f is forecasted biomass) from species distribution models (SDMs) for arrowtooth flounder (left) and walleye pollock (right) in the Bering Sea (select years between 1982 and 2018; see the Supporting information for the full time series). Covariates included in each model type: S (static SDM) – spatial variation, long-term mean bottom temperature and static habitat covariates; D3 (complex dynamic SDM) – spatial, temporal and spatiotemporal variation, location- and year-specific bottom temperature, interannual index of cold pool extent and static habitat covariates. Top: models fit from 1982 through 2002 and forecasted in 2003. Middle: models fit from 1982 through 2009 and forecasted in 2010. Bottom: models fit from 1982 through 2009 and forecasted in 2017. Blue illustrates where model forecasts underpredicted relative to observations and red shows where model forecasts overpredicted relative to observations. The black polygon denotes the northern Bering Sea, which was inconsistently sampled throughout the time series.

in prep.) because groundfishes are likely to redistribute based on the availability of important prey in addition to abiotic processes. Furthermore, over- and under-predictions may be reduced by parameterizing models such that they better account for non-stationary relationships between species and their environments (Litzow et al. 2018, Muhling et al. 2020).

Applications for fisheries management

At present, there are several management applications for hindcasting fine-scale population metrics. Spatially integrated stock assessments and other spatial management procedures, which often rely on temporally static species distributions (Link et al. 2011, Punt et al. 2013), would benefit from dynamic model hindcasts that help set the stage for changes in stock status. Population percentiles such as those estimated in this and other studies (Shotwell et al. 2022) can

be used to identify interannual variation in spatial structure, inform EFH consultations that reference the current state of the ecosystem and improve our understanding about the cumulative impacts of climate change. Multispecies extensions of our work would also serve as informative additions to reports that are designed to track community-level trends in ecosystem components (Siddon 2021).

For forecasting, it appears as though persistence forecasts from status quo static SDMs generate reasonable near-term forecasts. Forecasting experiments in other fields have come to a similar conclusion due to the lack of improvement in forecast skill from dynamic models relative to static baselines (Harris et al. 2018). Although continuing to improve forecasts from dynamic SDMs is warranted, especially in light of rapid climate change, static SDMs (or persistence forecasts from dynamic SDMs) may be more advantageous at the present moment. In addition to providing realistic forecasts,

static SDMs are less data intensive and generate shorter run times – making them more logistically feasible and therefore attractive to agencies with limited resources. We speculate, however, that continued development of dynamic SDMs will serve as an important waypoint in the advancement of high-resolution, integrated modeling used to provide climate-ready management advice (Hollowed et al. 2020).

Author contributions

Cheryl L. Barnes: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Project administration (equal); Resources (equal); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Timothy E. Essington**: Conceptualization (equal); Funding acquisition (equal); Methodology (supporting); Project administration (equal); Resources (equal); Supervision (equal); Writing – review and editing (supporting). **Jodi L. Pirtle**: Conceptualization (equal); Data curation (supporting); Funding acquisition (equal); Methodology (supporting); Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **Christopher N. Rooper**: Conceptualization (equal); Data curation (supporting); Methodology (supporting); Supervision (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Edward A. Laman**: Data curation (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Kirstin K. Holsman**: Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Supervision (supporting); Writing – review and editing (supporting). **Kerim Y. Aydin**: Conceptualization (supporting); Supervision (supporting); Writing – review and editing (supporting). **James T. Thorson**: Conceptualization (equal); Formal analysis (supporting); Funding acquisition (lead); Methodology (supporting); Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06189>.

Data availability statement

Input data and script files can be found at: <https://github.com/cheryl-barnes/SDM-fit-forecast> (Barnes et al. 2022).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Albouy, C. et al. 2014. From projected species distribution to food–web structure under climate change. – Global Change Biol. 20: 730–741.

Allyn, A. J. et al. 2020. Comparing and synthesizing quantitative distribution models and qualitative vulnerability assessments to project marine species distributions under climate change. – PLoS One 15: e0231595.

Barnes, C. L. et al. 2020. Development of a predation index to assess trophic stability in the Gulf of Alaska. – Ecol. Appl. 30: e02141.

Barnes, C. L. et al. 2022. Data from: Climate-informed models benefit hindcasting but present challenges when forecasting species–habitat associations. – Github Digital Repository, <https://github.com/cheryl-barnes/SDM-fit-forecast>.

Barnett, L. A. K. et al. 2021. Improving estimates of species distribution change by incorporating local trends. – Ecography 44: 427–439.

Berger, A. M. et al. 2017. Introduction to 'Space oddity: recent advances incorporating spatial processes in the fishery stock assessment and management interface'. – Can. J. Fish. Aquat. Sci. 74: 1693–1697.

Best, D. J. and Roberts, E. D. 1975. Algorithm AS 89: the upper tail probabilities of Spearman's rho. – Appl. Stat. 24: 377–379.

Birkmanis, C. A. et al. 2020. Future distribution of suitable habitat for pelagic sharks in Australia under climate change models. – Front. Mar. Sci. 7: 570.

Brodie, S. et al. 2020. Tradeoffs in covariate selection for species distribution models: a methodological comparison. – Ecography 43: 11–24.

Conn, P. B. et al. 2015. On extrapolating past the range of observed data when making statistical predictions in ecology. – PLoS One 10: e0141416.

Du Preez, C. and Tunnicliffe, V. 2011. Shortspine thornyhead and rockfish (Scorpaenidae) distribution in response to substratum, biogenic structures and trawling. – Mar. Ecol. Prog. Ser. 425: 217–231.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Grüss, A. et al. 2021. Synthesis of interannual variability in spatial demographic processes supports the strong influence of cold-pool extent on eastern Bering Sea walleye pollock *Gadus chalcogrammus*. – Progr. Ocean. 194: 102569.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Model. 135: 147–186.

Guisan, A. et al. 1999. GLM versus CCA spatial modeling of plant species distribution. – Plant Ecol. 143: 107–122.

Harris, D. J. et al. 2018. Forecasting biodiversity in breeding birds using best practices. – PeerJ 2018: 1–27.

Hauke, J. and Kossowski, T. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. – Quaest. Geogr. 30: 87–93.

Hazen, E. L. et al. 2018. A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. – Sci. Adv. 4: eaar3001.

Hermann, A. J. et al. 2013. A multivariate analysis of observed and modeled biophysical variability on the Bering Sea shelf: multi-decadal hindcasts (1970–2009) and forecasts (2010–2040). – Deep-Sea Res. II Top. Stud. Oceanogr. 94: 121–139.

Hermann, A. J. et al. 2019. Projected biophysical conditions of the Bering Sea to 2100 under multiple emission scenarios. – ICES J. Mar. Sci. 75: 1280–1304.

Hobday, A. J. et al. 2016. Seasonal forecasting for decision support in marine fisheries and aquaculture. – Fish. Oceanogr. 25: 45–56.

Hoff, G. R. 2016. Results of the 2016 eastern Bering Sea upper continental slope survey of groundfish and invertebrate resources. – US Dept Comm NOAA Tech Memo. NMFS AFSC-339, 272 pp.

Hollander, M. and Wolfe, D. A. 1973. Nonparametric statistical methods. – Wiley, pp. 185–194.

Hollowed, A. B. et al. 2020. Integrated modeling to evaluate climate change impacts on coupled social–ecological systems in Alaska. – Front. Mar. Sci. 6: 1–18.

Horn, B. K. P. 1981. Hill shading and the reflectance map. – Proc. IEEE 69: 14–47.

Hunsicker, M. E. et al. 2011. Functional responses and scaling in predator–prey interactions of marine fishes: contemporary issues and emerging concepts. – Ecol. Lett. 14: 1288–1299.

Huntington, H. P. et al. 2020. Evidence suggests potential transformation of the Pacific Arctic ecosystem is underway. – Nat. Clim. Change 10: 342–348.

Ianelli, J. et al. 2020. Assessment of the walleye pollock stock in the eastern Bering Sea. – NPFMC Bering Sea and Aleutian Islands SAFE: EBS Walleye Pollock, 173 pp.

Jurado-Molina, J. and Livingston, P. A. 2002. Climate-forcing effects on trophically linked groundfish populations: implications for fisheries management. – Can. J. Fish. Aquat. Sci. 59: 1941–1951.

Kearney, K. 2021. Temperature data from the eastern Bering Sea continental shelf bottom trawl surveys as used for hydrodynamic model validation and comparison. – NOAA Tech Mem 415, 48 pp.

Kearney, K. et al. 2020. A coupled pelagic–benthic–sympagic biogeochemical model for the Bering Sea: documentation and validation of the BESTNPZ model (v2019.08.23) within a high-resolution regional ocean model. – Geosci. Model Dev. 13: 597–650.

Kesner-Reyes, K. et al. 2020. AquaMaps: algorithm and data sources for marine organisms. – In: Froese, R. and Pauly, D. (eds), FishBase, v10/2019. <www.fishbase.org>.

Kleisner, K. M. et al. 2017. Marine species distribution shifts on the U.S. Northeast Continental Shelf under continued ocean warming. – Progr. Oceanogr. 153: 24–36.

Laman, E. A. et al. 2015. Correlating environmental and biogenic factors with abundance and distribution of Pacific ocean perch *Sebastes alutus* in the Aleutian Islands, Alaska. – Fish Bull. 113: 270–289.

Laman, E. A. et al. 2018. Using species distribution models to describe essential fish habitat in Alaska. – Can. J. Fish. Aquat. Sci. 75: 1230–1255.

Lauth, R. and Acuna, E. 2007. 2005 bottom trawl survey of the eastern Bering Sea continental shelf. – AFSC Proc. Rep. US Dept Comm. 2007-1, 164 pp.

Lauth, R. R. et al. 2019. Results of the 2017 eastern and northern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna. – NOAA Tech Mem 396.

Link, J. S. et al. 2011. Guidelines for incorporating fish distribution shifts into a fisheries management context. – Fish Fish. 12: 461–469.

Litzow, M. A. et al. 2018. Non-stationary climate–salmon relationships in the Gulf of Alaska. – Proc. R. Soc. B 285: 20181855.

Malecha, P. W. et al. 2005. Living substrate in Alaska: distribution, abundance and species associations. – In: Barnes, P. W. and Thomas, J. P. (eds), Benthic habitats and the effects of fishing. Am. Fish. Soc. Symp. 41. Am. Fish. Soc. Bethesda, MD, pp. 289–299.

Merow, C. et al. 2014. What do we gain from simplicity versus complexity in species distribution models? – Ecography 37: 1267–1281.

Mueter, F. J. and Litzow, A. M. 2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. – Ecol. Appl. 18: 309–320.

Muhling, B. A. et al. 2020. Predictability of species distributions deteriorates under novel environmental conditions in the California Current system. – Front. Mar. Sci. 7: 589.

National Marine Fisheries Service (NMFS) Alaska Region 2005. Final environmental impact statement for essential fish habitat identification and conservation in Alaska. – <https://repository.library.noaa.gov/view/noaa/17391>.

Norberg, A. et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. – Ecol. Monogr. 89: e10370.

Ovaskainen, O. and Soininen, J. 2011. Making more out of sparse data: hierarchical modeling of species communities. – Ecology 92: 289–295.

Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – Ecol. Lett. 2: 561–576.

Pebesma, E. 2018. Simple features for R: standardized support for spatial vector data. – R J. 10: 439–446.

Pebesma, E. J. 2004. Multivariate geostatistics in S: the gstat package. – Comp. Geosci. 30: 683–691.

Pennino, M. G. et al. 2020. Current and future influence of environmental factors on small pelagic fish distributions in the northwestern Mediterranean Sea. – Front. Mar. Sci. 7: 622.

Perry, A. L. et al. 2005. Climate change and distribution shifts in marine fishes. – Science 308: 1912–1915.

Pershing, A. J. et al. 2019. Challenges to natural and human communities from surprising ocean temperatures. – Proc. Natl Acad. Sci. USA 116: 18378–18383.

Pinsky, M. L. et al. 2018. Preparing ocean governance for species on the move. – Science 360: 1189–1191.

Pirtle, J. L. et al. 2019. Habitat suitability models for groundfish in the Gulf of Alaska. – Deep-Sea Res. Pt. II 165: 303–321.

Poloczanska, E. S. et al. 2013. Global imprint of climate change on marine life. – Nat. Clim. Change 3: 919–925.

Punt, A. E. et al. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. – ICES J. Mar. Sci. 70: 16–33.

Ready, J. et al. 2010. Predicting the distributions of marine organisms at the global scale. – Ecol. Model. 221: 467–478.

Rebekic, A. et al. 2015. Pearson's or Spearman's correlation coefficient – which one to use? – Agriculture 21: 47–54.

Richwine, K. A. et al. 2018. Surficial sediments of the eastern Bering Sea continental shelf: EBSSED-2 database documentation. – US Dept Comm, NOAA Tech Mem. NMFS-AFSC-377.

Roberts, D. R. et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical or phylogenetic structure. – Ecography 40: 913–929.

Rooper, C. N. et al. 2005. Using ecologically based relationships to predict distribution of flathead sole *Hippoglossoides elassodon* in the eastern Bering Sea. – Mar. Ecol. Prog. Ser. 290: 251–262.

Rooper, C. N. et al. 2014. Predictive models of coral and sponge distribution, abundance and diversity in bottom trawl surveys of the Aleutian Islands, Alaska. – Mar. Ecol. Prog. Ser. 503: 157–176.

Rooper, C. N. et al. 2016. Validation and improvement of species distribution models for structure-forming invertebrates in the eastern Bering Sea with an independent survey. – Mar. Ecol. Prog. Ser. 551: 117–130.

Rooper, C. N. et al. 2020. Predicted shifts of groundfish distribution in the Eastern Bering Sea under climate change, with implications for fish populations and fisheries management. – ICES J. Mar. Sci. 78: 220–234.

Saba, V. S. et al. 2016. Enhanced warming of the northwest Atlantic Ocean under climate change. – J. Geophys. Res. Oceans 121: 118–132.

Shotwell, S. K. et al. 2022. Synthesizing integrated ecosystem research to create informed stock-specific indicators for next generation stock assessments. – Deep-Sea Res. II Top. Stud. Oceangr. 198: 105070.

Sibson, R. 1981. A brief description of natural neighbor interpolation. – In: Barnet, V. (ed.), Interpolating multivariate data. Wiley, pp. 21–36.

Siddon, E. 2021. Ecosystem Status Report 2021: Eastern Bering Sea. – NPFMC Bering Sea and Aleutian Islands SAFE, 249 pp.

Sigler, M. F. et al. 2015. Faunal features of submarine canyons on the eastern Bering Sea slope. – Mar. Ecol. Prog. Ser. 526: 21–40.

Silva, C. R. et al. 2018. The role of Pleistocene climate change in the genetic variability, distribution and demography of *Proechimys cuvieri* and *P. guyannensis* (Rodentia: Echimyidae) in northeastern Amazonia. – PLoS One 13: e0206660.

Stabeno, P. J. et al. 2001. On the temporal variability of the physical environment over the south-eastern Bering Sea. – Fish. Oceanogr. 10: 81–98.

Stabeno, P. J. et al. 2012. Comparison of warm and cold years on the southeastern Bering Sea shelf and some implications for the ecosystem. – Deep-Sea Res. II 65–70: 31–45.

Stabeno, P. J. et al. 2017. Return of warm conditions in the southeastern Bering Sea: physics to fluorescence. – PLoS One 12: e0185464.

Stahl, J. P. and Kruse, H. G. 2008. Spatial and temporal variability in size at maturity of walleye pollock in the eastern Bering Sea. – Trans. Am. Fish. Soc. 137: 1543–1557.

Stark, J. W. 2012. Female maturity, reproductive potential, relative distribution, and growth compared between arrowtooth flounder (*Atheresthes stomias*) and Kamchatka flounder (*A. evermanni*) indicating concerns for management. – J. Appl. Ichthyol. 28: 226–230.

Stevenson, D. E. and Lauth, R. R. 2019. Bottom trawl surveys in the northern Bering Sea indicate recent shifts in the distribution of marine species. – Polar Biol. 42: 407–421.

Stone, R. et al. 2011. A guide to the deep-water sponges of the Aleutian Island archipelago. – NOAA Professional Paper NMFS 12, 187 pp.

Thorson, J. T. 2019a. 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. 2019b. 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. 64: 2632–2645.

Thorson, J. T. et al. 2016. Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. – Methods Ecol. Evol. 7: 990–1002.

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.

Venables, W. N. and Ripley, D. B. 2002. Modern applied statistics with S, 4th edn. – Springer.

Walbridge, S. et al. 2018. Unified geomorphological analysis workflows with Benthic Terrain Modeler, v3. – Geosciences 8: 94.

Wang, M. et al. 2012. Future climate of the Bering and Chukchi Seas projected by global climate models. – Deep-Sea Res. II 65–70: 46–57.

Wilderbuer, T. K. et al. 2010. Kamchatka flounder. – NPFMC Bering Sea and Aleutian Islands SAFE, pp. 763–780.

Wood, S. N. 2003. Thin-plate regression splines. – J. R. Stat. Soc. 65: 95–114.

Wood, S. N. 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. – J. Am. Stat. Assoc. 99: 673–686.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. – J. R. Stat. Soc. 73: 3–36.

Wood, S. N. 2017. Generalized additive models: an introduction with R, 2nd edn. – Chapman and Hall/CRC.

Wood, S. N. et al. 2016. Smoothing parameter and model selection for general smooth models (with discussion). – J. Am. Stat. Assoc. 111: 1548–1575.

Zimmermann, M. and Prescott, M. 2018. Bathymetry and canyons of the eastern Bering Sea slope. – Geosciences 8: 184.

Zimmermann, M. et al. 2013. Smooth sheet bathymetry of the Aleutian Islands. – US Dept Comm NOAA Tech Memo. NMFS-AFSC-250, 43 pp.

Zurell, D. et al. 2020. A standard protocol for reporting species distribution models. – Ecography 43: 1261–1277.