

# Forecast skill for predicting distribution shifts: A retrospective experiment for marine fishes in the Eastern Bering Sea

James T. Thorson 

Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington

#### Correspondence

James T. Thorson, Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA.  
Email: [james.thorson@noaa.gov](mailto:james.thorson@noaa.gov)

#### Present address

James T. Thorson, Habitat and Ecosystem Process Research Program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington

## Abstract

Forecasting distribution shifts under novel environmental conditions is a major task for ecologists and conservationists. Researchers forecast distribution shifts using several tools including: predicting from an empirical relationship between a summary of distribution (population centroid) and annual time series ("annual regression," AR); or fitting a habitat-envelope model to historical distribution and forecasting given predictions of future environmental conditions ("habitat envelope," HE). However, surprisingly little research has estimated forecast skill by fitting to historical data, forecasting distribution shifts and comparing forecasts with subsequent observations of distribution shifts. I demonstrate the important role of retrospective skill testing by forecasting poleward movement over 1-, 2- or 3-year periods for 20 fish and crab species in the Eastern Bering Sea and comparing forecasts with observed shifts. I specifically introduce an alternative vector-autoregressive spatio-temporal (VAST) forecasting model, which can include species temperature responses, and compare skill for AR, HE and VAST forecasts. Results show that the HE forecast has 30%–43% greater variance than predicting that future distribution is identical to the estimated distribution in the final year (a "persistence" forecast). Meanwhile, the AR explains 2%–6% and VAST explains 8%–25% of variance in poleward movement, and both have better performance than a persistence forecast. HE and AR both generate forecast intervals that are too narrow, while VAST models with or without temperature have appropriate width for forecast intervals. Retrospective skill testing for more regions and taxa should be used as a test bed to guide future improvements in methods for forecasting distribution shifts.

#### KEY WORDS

distribution shift, forecast skill, habitat-envelope model, short-term forecast, spatio-temporal model

## 1 | INTRODUCTION

Many ecosystems worldwide are showing rapid responses to changing environmental conditions. Distribution shifts often indicate the intensity of environmental impacts, both because distribution shifts can directly cause changes in community stability (Theobald, Breckheimer, & HilleRisLambers, 2017) and because

distribution is an integrated measure of changing demographic rates (Merow et al., 2014). Rapid distribution shifts have been documented for birds (Hitch & Leberg, 2007), plants (Kelly & Goulden, 2008) and fishes (Pinsky, Worm, Fogarty, Sarmiento, & Levin, 2013).

Distribution shifts due to environmental change can be particularly rapid in marine ecosystems. For example, the Gulf of Maine has

shown rapid increases in water temperature during the past decade, and these increases have changed the spatial distribution and phenology of lobster fisheries thereby causing economic hardships in 2012 (Mills et al., 2013). Similarly, the Northwest Atlantic as a whole has shown rapid warming, and this has driven a rapid northward shift in abundance for yellowtail and summer flounder, in turn causing a lagged shift in fishing effort for these species (Pinsky & Fogarty, 2012). Rapid changes such as these are less frequent in ecological systems where slow changes in habitat or long-lived life-history strategies can underlie a lagged effect of climate on local population densities (Schurr et al., 2012).

Forecasting future shifts in distribution is a unifying challenge for ecologists and conservationists across many taxa. Forecasting future distribution shift is important to predict likely changes in competitive and facultative interactions among species (Schliep et al., 2018) or the invasion success of newly introduced species (Ramírez-Albores, Bustamante, & Badano, 2016). In marine species, forecasting distribution shifts is particularly important when renegotiating fishery access rights (Pinsky & Fogarty, 2012), or when forecasting potential impacts of changing climate on fishery economic potential (Cheung et al., 2010).

Recent research has recommended forecasting distribution shifts using mechanistic models that represent the impact of environmental changes and biotic interactions on individual demographic rates (Mellin et al., 2016; Swab, Regan, Matthies, Becker, & Bruun, 2015; Trainor, Schmitz, Ivan, & Shenk, 2014; Zurell, 2017). However, many taxa (e.g., marine fishes, insects and fungi communities, among others) often do not have sufficient individual-specific demographic information to parameterize stage-based models of environmental and biological mechanisms (e.g., examples in Parmesan, 2006). In marine fishes, for example, researchers have instead forecasted distribution shifts using two types of models: (a) annual regressions (AR) of distribution shift, or (b) habitat-envelope (HE) models. Annual regressions of distribution generally involve calculating an annual summary statistic for species distribution, for example, the centroid of a population in each year. This centroid is then regressed against multiple annual covariates, for example, total abundance, average bottom temperature or fishing mortality rates (Adams et al., 2018; Mueter & Litzow, 2008; Nye, Link, Hare, & Overholtz, 2009; Spencer, 2008), and this empirical relationship is used to forecast distribution shift under alternative future values of these covariates (Hare, Alexander, Fogarty, Williams, & Scott, 2010). By contrast, habitat-envelope models involve estimating or pre-specifying the habitat suitability function for a given species based on temperature, demersal habitat or other local habitat conditions (Cheung, Lam, & Pauly, 2008; Kaschner, Watson, Trites, & Pauly, 2006). Changes in habitat suitability can then be forecasted given projected changes in environmental conditions, and changed habitat suitability can be used to forecast changes in the centroid of a population's distribution (Cheung et al., 2009). Recent research has also developed spatio-temporal models to estimate distribution shifts and attribute shifts to different causal mechanisms, and these models can include both annual covariates (e.g., average environmental

1 INTRODUCTION	159
2 METHODS	161
2.1 Annual-regression forecasts	161
2.2 Habitat-envelope forecasts	161
2.3 Vector-autoregressive spatio-temporal (VAST) forecasts	162
2.4 Retrospective database: Marine species in the Eastern Bering Sea	163
2.5 Metrics of forecast skill	163
2.5.1 Performance relative to persistence forecast	163
2.5.2 Usefulness of uncertainty estimates	164
3 RESULTS	164
3.1 Descriptive results across species	164
3.2 Illustration of forecast models using an example, <i>Gadus chalcogrammus</i>	165
3.3 Comparison of forecast skill across all 20 species	166
4 DISCUSSION	167
ACKNOWLEDGEMENTS	171
DATA ACCESSIBILITY	171
REFERENCES	171
SUPPORTING INFORMATION	173

conditions) and local covariates affecting habitat suitability (Thorson, Ianelli, & Kotwicki, 2017; Tredennick et al., 2016).

Despite this increased interest in forecasting species distribution, there has been relatively little research comparing predictive skill for alternative potential models used to forecast distribution shifts. Skill testing is a common approach to compare models and benchmark progress in many other types of environmental forecasting and has been widely used, for example, to validate strategic forecasts of abundance and productivity using an ecosystem model for the Northwest Atlantic (Olsen et al., 2016), explore improvements in hurricane forecasts over time (Vitart, 2014), assess skill for decadal forecasts of butterfly distribution in Finland (Eskildsen et al., 2013), or identify when regression models are useful to forecast fish distribution for the Northwest Atlantic (Kleisner et al., 2017). Skill testing has seen increasing use for seasonal forecasts of fish distribution shift, for example, 2-month forecasts for the spatial distribution of tunas in the Great Australian Bight (Eveson, Hobday, Hartog, Spillman, & Rough, 2015), or 4- to 8-month forecasts of sardine distribution in the US California Current (Kaplan, Williams, Bond, Hermann, & Siedlecki, 2016). However, skill testing has not been widely applied to compare performance for multiple models to forecast shifts in species distribution occurring over time-scales between seasonal and climate forecasts (1–5 years, termed "short-term" in the following).

In this paper, I use retrospective skill testing to compare the performance of "annual regression" (AR) and "habitat-envelope" (HE) forecasts of distribution shift with an alternative vector-autoregressive

spatio-temporal (VAST) model. Following common practice in fisheries, I measure distribution as the centroid of population biomass rather than measuring changes in occupied habitat. This VAST and HE models involve fitting a delta-generalized linear mixed model (delta-GLMM) to local samples of population biomass while accounting for both spatial, temporal and spatio-temporal correlations in biomass density. AR, HE and VAST forecasts are generated by fitting to biomass-sampling data at 370 stations for each of 20 marine species in the Eastern Bering Sea from 1982 to 2015. For each species, I fit data for a subset of years, predict shifts in centre of gravity (COG) occurring after the last fitted year and compare the prediction with observed shifts over those subsequent years. Finally, I summarize predictive skill by calculating the variance explained by each model's forecast relative to forecasting that the population does not move in the future (termed a "persistence forecast"). I also determine whether the predicted uncertainty for each model (termed the "forecast interval") is too narrow, too wide or has appropriate width. These two metrics of predictive skill measure whether forecasts are accurate and generate useful estimates of forecast uncertainty. I use this example to illustrate the importance of retrospective skill testing for measuring model performance when planning future model developments or choosing among alternative forecast models.

## 2 | METHODS

I seek to evaluate predictive skill for three potential models for forecasting distribution shifts without invoking detailed demographic mechanism: annual-regression models (AR), habitat-envelope models (HE) and a vector-autoregressive spatio-temporal (VAST) model with or without temperature as a habitat covariate. These models forecast distribution based on changes in temperature, and I here assume that models have perfect information about future temperature; results therefore represent an "optimistic" picture of forecast skill relative to cases when future temperature must itself be forecasted. I first discuss each of these in detail (see Table 1 for summary). I then describe two measures of forecast skill: reduction in mean squared error (MSE) relative to a null "persistence" forecast,

and predictive interval coverage. Finally, I describe the data set used in this skill testing experiment.

### 2.1 | Annual-regression forecasts

Several authors have previously analysed environmental impacts on distribution shifts by regressing a measure of population location against annual covariates, without proposing any explicit model for how the environment impacts local population densities (Hare et al., 2010; Nye et al., 2009; Pinsky et al., 2013). This approach will presumably perform better than the habitat-envelope or spatio-temporal models whenever the impact of habitat covariates on local densities is difficult to specify correctly a priori.

To implement this approach, I first develop a statistic  $Y(t)$  representing the centroid of the population north of the equator, calculated as the "abundance-and-area weighted average" (AAWA) of sampled locations (see Supporting Information Appendix S1). I then regress  $Y(t)$  against average bottom temperature  $\bar{C}(t)$  in year  $t$ :

$$Y(t) \sim \text{Normal}(\alpha + \delta \bar{C}(t), \sigma_Y^2) \quad (1)$$

where  $\alpha$  and  $\delta$  represent the intercept and slope for a linear regression of  $Y(t)$  on average temperature  $\bar{C}(t)$ . Finally, I use estimated parameters from this linear regression to forecast future values of  $Y(t_{\text{forecast}})$  given  $\bar{C}Y(t_{\text{forecast}})$  as well as its standard errors.

### 2.2 | Habitat-envelope (HE) forecasts

Habitat-envelope (a.k.a. species distribution/density) models typically use available data or expert opinion to define habitat suitability as a function of habitat variables, where future distribution can then be forecasted given alternative values for habitat variables (Cheung et al., 2009; Eskildsen et al., 2013; Eveson et al., 2015).

I use a delta-generalized linear model (Lo, Jacobson, & Squire, 1992) to separately predict occupied habitat via changes in encounter probability  $p$  as well as biomass density within occupied habitat via positive catch rates  $r$ . This delta model involves specifying a probability distribution for sampled biomass  $b_i$  for each sample  $i$  given predicted encounter probability  $p_i$  and positive catch rates  $r_i$ :

**TABLE 1** A description of forecasting models used in this study including what data they fit, which covariates they use, whether they account for residual spatial patterns in biomass density, and what information they use to inform future forecasts of movement

Model name	Data used	Covariates used	Accounts for residual spatial patterns?	Type of information regarding future distribution
Habitat-envelope model	Field samples of population biomass	Spatially explicit bottom temperature	No	Temperature landscape in each year
Vector-autoregressive spatio-temporal model (VAST)	Field samples of population biomass	None.	Yes	Persistence of spatio-temporal hotspots
VAST model with temperature	Field samples of population biomass	Spatially explicit bottom temperature	Yes	Temperature landscape in each year, and persistence of spatio-temporal hotspots
Annual regression	Abundance-and-area weighted average	Average bottom temperature in each year	No	Annual increase/decrease in average temperature across landscape

$$\Pr(b_i = B) = \begin{cases} 1-p_i & \text{if } B = 0 \\ p_i \times \text{Lognormal}(B; \log(r_i), \sigma_b^2) & \text{if } B > 0 \end{cases} \quad (2)$$

Specifically, I use a Poisson-link delta model (Thorson, 2018) that accomplishes this by modelling both the density of individuals  $n(s, t)$  and the average biomass per individual  $w(s, t)$  at each location  $s$  and year  $t$ . Encounter probability  $p_i$  is defined in the Poisson-link delta model given the assumption that individuals are randomly distributed in the vicinity of sampling, such that the probability of encountering at least one individual follows:

$$p_i = 1 - \exp(-\alpha_i \times n(s_i, t_i)) \quad (3)$$

including an offset for the area  $a_i$  sampled for sample  $i$  (in this case, the area swept by each operation of bottom trawl sampling). Positive catch rate  $r_i$  is defined from the definition of biomass density,  $d(s, t) = r(s, t)p(s, t) = n(s, t)w(s, t)$ :

$$r_i = \frac{a_i \times n(s_i, t_i)}{p_i} \times w(s_i, t_i) \quad (4)$$

Finally, I define numbers density and average weight as quadratic functions of bottom temperature using a log-link for each:

$$\begin{aligned} \log(n(s, t)) &= \beta_n(t) + \gamma_{n1}C(s, t) + \gamma_{n2}C(s, t)^2 \\ \log(w(s, t)) &= \beta_w(t) + \gamma_{w1}C(s, t) + \gamma_{w2}C(s, t)^2 \end{aligned} \quad (5)$$

where  $C(s, t)$  is temperature (in degrees Celsius) at each location  $s$  and year  $t$ , and  $\gamma_{n1}, \gamma_{n2}, \gamma_{w1}$  and  $\gamma_{w2}$  represent the estimated log-linear and log-quadratic impact of temperature on numbers density and average weight. I model annual intercepts  $\beta_n(t)$  and  $\beta_w(t)$  as a random-walk process:

$$\begin{aligned} \beta_n(t+1) &\sim \text{Normal}(\beta_n(t), \sigma_n^2) \\ \beta_w(t+1) &\sim \text{Normal}(\beta_w(t), \sigma_w^2) \end{aligned} \quad (6)$$

where the variance of this random-walk process,  $\sigma_n^2$  and  $\sigma_w^2$ , is estimated. This Poisson-link delta model implies that temperature has a quadratic effect on predicted log-density:

$$\log(d(s, t)) \propto (\gamma_{n1} + \gamma_{w1})C(s, t) + (\gamma_{n2} + \gamma_{w2})C(s, t)^2 \quad (7)$$

where using the Poisson-link delta model ensures that the effect of changing temperature is additive between both components of the delta model (Thorson, 2018). Fixed effects ( $\gamma_{n1}, \gamma_{w1}, \gamma_{n2}, \gamma_{w2}, \sigma_n^2, \sigma_w^2$ ) are estimated by minimizing the negative log-likelihood function, while intercepts ( $\beta_n, \beta_w$ ) are treated as random effects.

This model is fitted to data for all years  $t_{\text{fitted}} \in \{t_{\text{initial}}, \dots, t_{\text{final}}\}$ , and then estimated parameters are used to predict density  $d(s, t)$  for all locations during forecast years  $t_{\text{forecast}} \in \{t_{\text{final}+1}, t_{\text{final}+2}, t_{\text{final}+3}\}$  given known bottom temperature  $C(s, t_{\text{forecast}})$  during forecast years. Finally, density predictions are used to calculate the centroid of the population's distribution,  $Z(t)$ :

$$Z(t) = \frac{\sum_{s=1}^{n_s} (Z(s) \times a(s) \times d(s, t))}{\sum_{s=1}^{n_s} (a(s) \times d(s, t))} \quad (8)$$

where  $a(s)$  is the area associated with each modelled location  $s$ , and  $Z(s)$  is a measure of location. In the following, I predict poleward movement and define  $z(s)$  as the distance of location  $s$  from the equator (in kilometres).

## 2.3 | Vector-autoregressive spatio-temporal (VAST) forecasts

This habitat-envelope model attributes all variation in spatial distribution to a quadratic effect of bottom temperature on log-density and neglects any residual spatial correlation. However, extensive previous research suggests that predictive performance and statistical interpretation are degraded when spatial models neglect spatial autocorrelation in residuals (Bahn & McGill, 2007; Dormann et al., 2007; Thorson, Ianelli, et al., 2017). I therefore apply an alternative spatio-temporal model that can incorporate quadratic temperature effects as well as residual spatial patterns (spatial patterns that are constant among all years) and spatio-temporal patterns (spatial patterns that vary among years). This "semiparametric" model has a habitat-envelope model as its deterministic skeleton, but is also able to identify areas with higher or lower density than expected, and to estimate how quickly these areas with higher/lower density revert to their expected density when forecasting forward in time (Kai, Thorson, Piner, & Maunder, 2017; Thorson, Ianelli, et al., 2016).

Specifically, I apply a vector-autoregressive spatio-temporal (VAST) model that has been used extensively elsewhere (Thorson, 2019; Thorson & Barnett, 2017), but which has not been used for short-term forecasts of spatial distribution. This model again uses a Poisson-link delta model (Equations 2–4) and a random-walk process for annual intercepts (Equation 6). However, the spatio-temporal model involves changing the linear predictor for log-numbers density  $n(s, t)$  and log-average weight  $w(s, t)$ :

$$\begin{aligned} \log(n(s, t)) &= \beta_n(t) + \omega_n(s) + \epsilon_n(s, t) + \gamma_{n1}C(s, t) + \gamma_{n2}C(s, t)^2 \\ \log(w(s, t)) &= \beta_w(t) + \omega_w(s) + \epsilon_w(s, t) + \gamma_{w1}C(s, t) + \gamma_{w2}C(s, t)^2 \end{aligned} \quad (9)$$

where  $\omega_n(s)$  and  $\omega_w(s)$  represent spatial variation, and  $\epsilon_n(s, t)$  and  $\epsilon_w(s, t)$  represent spatio-temporal variation in  $n$  and  $w$ . Spatial errors follow a Gaussian random field and are treated as random effects:

$$\begin{aligned} \omega_n &\sim \text{MVN}(0, \sigma_{\omega n}^2 \mathbf{R}_n) \\ \omega_w &\sim \text{MVN}(0, \sigma_{\omega w}^2 \mathbf{R}_w) \end{aligned} \quad (10)$$

where spatial correlations  $\mathbf{R}_n$  and  $\mathbf{R}_w$  follow a Matérn correlation function where I estimate geometric anisotropy and a separate decorrelation rate for  $\mathbf{R}_n$  and  $\mathbf{R}_w$  (see Supporting Information Appendix S2 for details). I use a "predictive process" approximation to simplify computation for spatial and spatio-temporal variation at 100 "knots" in a stochastic partial differential equation approximation to the Matérn correlation function (Lindgren, Rue, & Lindström, 2011). Spatio-temporal variation is specified similarly, but also follows an autoregressive process across years:

$$\begin{aligned} \epsilon_n(t+1) &\sim \text{MVN}(\rho_n \epsilon_n(t), \sigma_{\epsilon n}^2 \mathbf{R}_n) \\ \epsilon_w(t+1) &\sim \text{MVN}(\rho_w \epsilon_w(t), \sigma_{\epsilon w}^2 \mathbf{R}_w) \end{aligned} \quad (11)$$

where  $\rho_n$  and  $\rho_w$  are estimated parameters that govern the rate at which areas with higher/lower density than expected revert to the average spatial distribution during the forecast period.

I specifically explore two versions of this VAST model. The first version excludes temperature effects (i.e.,  $\gamma_{n1} = \gamma_{n2} = \gamma_{w1} = \gamma_{w2} = 0$ ), while the second estimates temperature effects; the two models are otherwise identical. The former model explains differences in biomass density  $d(s, t)$  purely via spatial and spatio-temporal residual terms, and I therefore refer to it as a “non-parametric” forecast model. The latter includes both non-parametric components and a parametric effect of temperature. I therefore follow previous studies in calling it a “semiparametric” forecast model (Shelton, Thorson, Ward, & Feist, 2014; Sugeno & Munch, 2013; Thorson, Ono, & Munch, 2014). I fit this model using package VAST (<https://github.com/James-Thorson/VAST>) in the R statistical environment (R Core Team 2016). VAST uses Template Model Builder to implement the Laplace approximation to the marginal likelihood of fixed effects (Kristensen, Nielsen, Berg, Skaug, & Bell, 2016), and a gradient-based non-linear optimizer to identify the value of fixed effects that maximizes this value. After the non-linear optimizer has converged, I then use a Newton optimizer to ensure that the gradient of the approximated marginal log-likelihood with respect to every fixed effect is  $<10^{-6}$ . This final step is implemented to ensure that the function maximizer is very tightly converged. I then confirm that parameters are estimable by confirming that the Hessian matrix is positive definite at the maximum-likelihood estimates. Parameter estimation using VAST takes approximately 10–20 min for each species and retrospective run, so the analysis was feasible without using of high-performance computing tools.

## 2.4 | Retrospective database: Marine species in the Eastern Bering Sea

I have defined four methods to forecast poleward movement of fish populations given available sampling data: habitat-envelope models, a spatio-temporal model with or without temperature, and an annual-regression estimator. I compare performance of these four estimators when fitting to data for the 20 most abundant (in terms of sampled biomass) fish and crab (decapod) species in the bottom trawl survey operated by the Alaska Fishery Science Center in the continental shelf of the Eastern Bering Sea from 1982 to 2015. This annual survey has used a fixed-station design with over 370 samples per year and consistent gear over this period (Lauth & Conner, 2016), and I download data from the AFSC website ([http://www.afsc.noaa.gov/RACE/groundfish/survey\\_data/data.htm](http://www.afsc.noaa.gov/RACE/groundfish/survey_data/data.htm)) using R package *FishData* (<https://github.com/James-Thorson/FishData>). Although this case-study involves a (nearly) identical sampling design across years, future research could apply the same method to data sets that have different sampling intensity or design among years (e.g., Thorson, Ianelli, et al., 2016).

For each species, I then fit each forecast model 21 times, that is, using data for all years 1982–2015, using data 1982–2014 and predicting 2015, using data 1982–2013 and predicting 2014–2015, ..., and using data for 1982–1995 and predicting 1996–2015. This

involves fitting four estimators in 21 retrospective models using data for 20 species (1,680 model fits total). For each model fit, I then record forecasts of the population centroid 1, 2 or 3 years ahead of the last fitted year. I also record the predicted standard error for this forecast, where asymptotic standard errors are calculated using a generalization of the delta method for empirical Bayes models (Kass & Steffey, 1989).

## 2.5 | Metrics of forecast skill

I follow Tommasi et al. (2017) in arguing that a well-performing estimator will have two characteristics: (a) It will outperform a “persistence” forecast; and (b) it will estimate uncertainty in a useful manner.

### 2.5.1 | Performance relative to persistence forecast

I evaluate each forecast model by comparing performance when predicting future changes in COG against the observed change in COG calculated using the abundance-and-area weighted average (AAWA) estimator. I evaluate skill when forecasting future changes in distribution because distribution shifts are likely to be disruptive to place-based management measures and fishing activities. I do not know the true centroid of the population in any year, but the AAWA estimator is independent of each forecast model and therefore represents a fair and unbiased metric against which to compare each forecast model. I specifically calculate changes in AAWA,  $\Delta Y_{AAWA} (\Delta t|t_{final})$ , when forecasting 1, 2 or 3 years forward from year  $t_{final}$ :

$$\Delta Y_{AAWA} (\Delta t|t_{final}) = Y(t_{final} + \Delta t) - Y(t_{final}) \quad (12)$$

where  $\Delta t \in \{1, 2, 3\}$ . I also calculate change in centroid  $\Delta Y(\Delta t|t_{final})$  over  $\Delta t$  forecast years for each estimator when fitting to data through year  $t_{final}$ . For example, given the forecasted centroid  $Y_{CE}(t_{forecast}|t_{final})$  in year  $t_{forecast}$  using data through  $t_{final}$  and the habitat-envelope model:

$$\Delta Y_{CE} (\Delta t|t_{final}) = Y_{CE}(t_{final} + \Delta t|t_{final}) - Y_{CE}(t_{final}|t_{final}) \quad (13)$$

The persistence forecast will predict that  $Y_{persistence}(t_{final} + \Delta t) = Y_{persistence}(t_{final})$  and therefore that  $\Delta Y_{persistence} (\Delta t) = 0$ . I therefore calculate the predictive error for each estimator relative to the persistence forecast as:

$$E(\Delta t|t_{final}) = \frac{\Delta Y_{AAWA} (\Delta t|t_{final}) - \Delta Y(\Delta t|t_{final})}{\Delta Y_{AAWA} (\Delta t|t_{final})} \quad (14)$$

where this formula is appropriate given preliminary analysis showing that the AAWA always shows some movement among years (i.e., the denominator  $\Delta Y_{AAWA} (\Delta t|t_{final})$  is never zero). I then calculate the mean squared error relative to the persistence forecast:

$$V(\Delta t) = \frac{1}{21 - \Delta t} \sum_{t_{final}=1995}^{2015 - \Delta t} E(\Delta t|t_{final})^2 \quad (15)$$

and I define the variance explained as  $R^2(\Delta t) = 1 - V(\Delta t)$ . A model performing as well as the persistence forecast will have  $V(\Delta t) = 1$  and  $R^2(\Delta t) = 0$ , while a model with  $R^2(\Delta t) > 0$  outperforms the persistence forecast while a model with  $R^2(\Delta t) < 0$  has degraded performance relative to a persistence forecast. Hypothetically, a model with, for example,  $R^2(\Delta t) = 0.2$  has explained 20% of the variance in future changes in the observed abundance-and-area weighted centroid.

### 2.5.2 | Usefulness of uncertainty estimates

I also evaluate model performance by asking which model can accurately estimate the uncertainty of their forecasts. I therefore calculate the quantile  $Q(\Delta t|t_{\text{final}})$  of the observed change  $\Delta Y_{\text{AWA}}$  ( $\Delta t|t_{\text{final}}$ ) given the predictive distribution,  $\Delta Y(\Delta t|t_{\text{final}})$  and  $SE(Y(t_{\text{final}} + \Delta t))$ :

$$Q(\Delta t|t_{\text{final}}) = \Phi(\Delta Y_{\text{AWA}}(\Delta t|t_{\text{final}}); \Delta Y(\Delta t|t_{\text{final}}), SE(Y(t_{\text{final}} + \Delta t))) \quad (16)$$

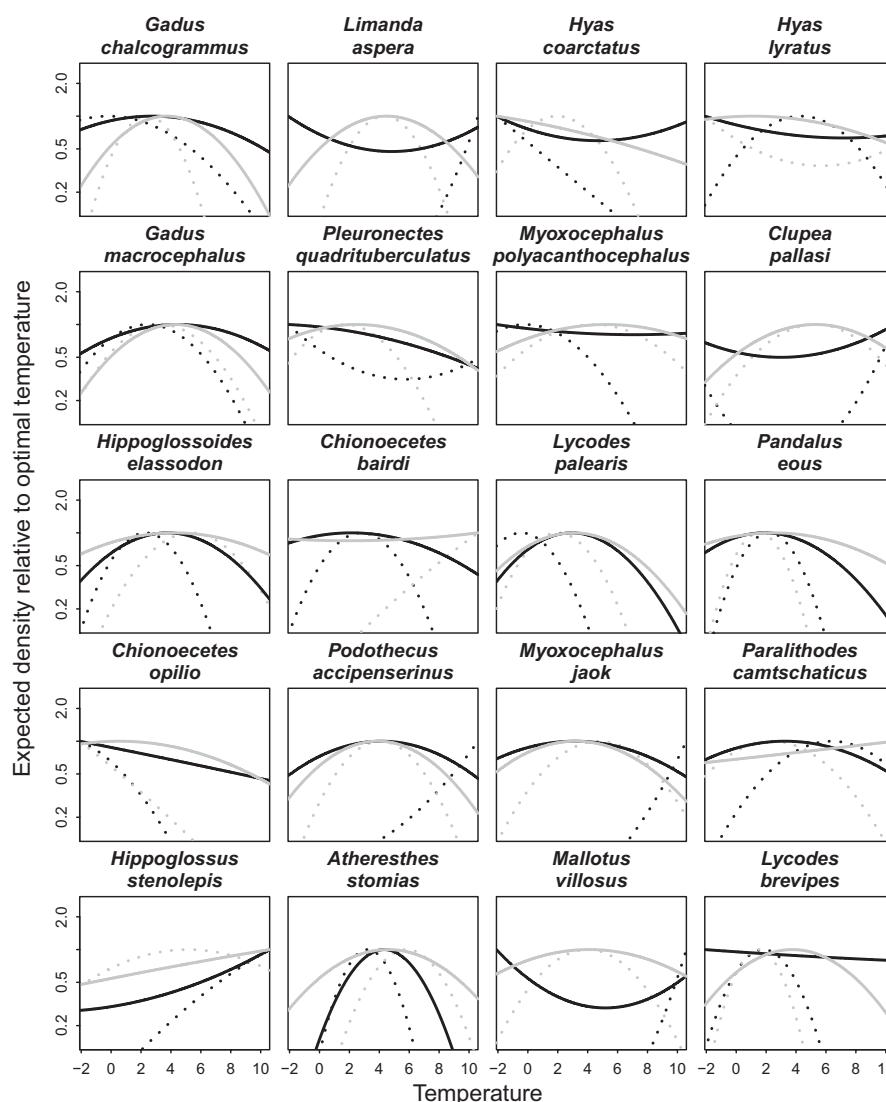
where  $\Phi(x; \mu, \sigma)$  is the cumulative distribution function for a normal distribution with mean  $\mu$  and standard deviation  $\sigma$  evaluated at location  $x$ , and where the forecast standard error  $SE(Y(t_{\text{final}} + \Delta t))$  is calculated using the delta method. For a well-performing model, this

quantile  $Q(\Delta t|t_{\text{final}})$  will be approximately uniform, while a model that underestimates its own uncertainty will have excess  $Q(\Delta t|t_{\text{final}})$  near 0 or 1, and a model that overestimates its own uncertainty will have excess  $Q(\Delta t|t_{\text{final}})$  near 0.5.

## 3 | RESULTS

### 3.1 | Descriptive results across species

Results show that the majority of species show an optimal temperature within the range observed in the Eastern Bering Sea, such that changes in temperature could drive changes in spatial preferences. Specifically, the habitat-envelope model and VAST model with temperature both involve a quadratic effect of temperature on log-biomass density. Comparing across species, both models for the majority of species show decreasing biomass density as temperature goes to either the highest or lowest values observed in the Eastern Bering Sea (Figure 1). Specifically, there are 40 estimated “temperature responses” (affecting numbers density and average biomass for each of 20 species) for each model estimating



**FIGURE 1** Temperature responses for each species (panels) estimated using the habitat-envelope model (dotted lines) or the VAST model with temperature (solid lines). Each panel shows the predicted density (y-axis) for a given temperature (x-axis) relative to the optimal temperature for a given species among the range of temperatures observed within the Eastern Bering Sea, showing either the predicted response of group density  $n$  to temperature (black lines) or the predicted response of biomass per group  $w$  to temperature (grey lines)

a quadratic effect of temperature, and a negative curvature is observed in 28 temperature effects for the habitat-envelope model and 26 effects for the VAST model with temperature (see Figure 1). The VAST model often estimates a more broad temperature tolerance than the habitat-envelope model, for example, for arrowtooth flounder (*Atherestes stomias*, Pleuronectidae) where the temperature response has nearly twice the width for the VAST model than the habitat-envelope model for both numbers density and average-weight components of the delta model.

The VAST model with temperature also shows positive auto-correlation among years for spatio-temporal residuals for all species (median across species:  $\rho_n = 0.86$  and  $\rho_w = 0.65$ , Table 2). This indicates that areas with higher/lower density than expected are likely to exhibit higher/lower density for several years sequentially. As a consequence, any year with a centre of gravity that is more skewed towards the pole/equator than its average value will often be followed by another year with a distribution similarly skewed towards the pole/equator.

### 3.2 | Illustration of forecast models using an example, *Gadus chalcogrammus*

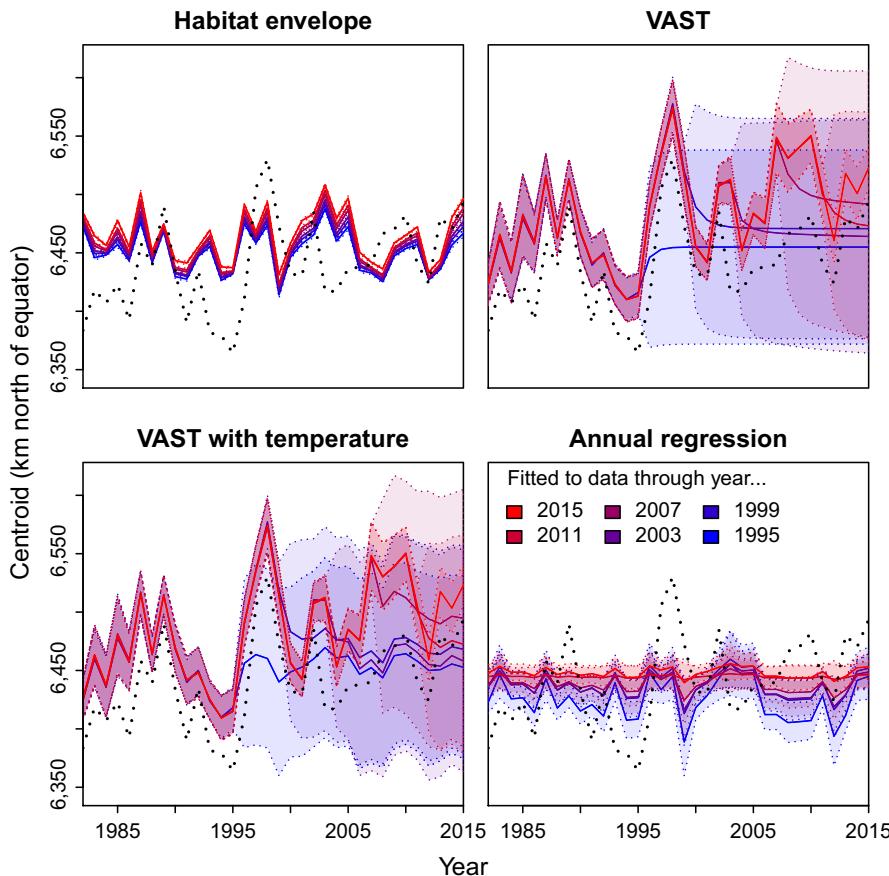
Comparing forecasts of distribution for Alaska pollock (*Gadus chalcogrammus*, Gadidae) using all four estimators shows that the

habitat-envelope and annual-regression estimators capture neither the northward shift in the abundance-and-area weighted average (AAWA) estimator nor the large interannual variation in this estimator (Figure 2). For example, neither HE nor AR estimators capture the nearly 200 km poleward movement from 1995 to 1998. By contrast, the spatio-temporal estimators with or without temperature show similar performance, and both capture the observed northward shift for this species when fitted to data for 1982–2015. Both also generate forecast intervals that are much wider than the habitat-envelope and annual-regression estimators, and these wide forecast intervals appear to be necessary to contain estimates of centre of gravity arising when fitting to subsequent data. For example, the spatio-temporal estimator with temperature using data for 1982–1995 predicts a rapid increase in confidence interval width when predicting COG for 1996–1998. This rapid increase in confidence interval width is appropriate, given that 1998 has the most northward distribution during the entire period 1982–2015 and is still somewhat outside the estimated forecast interval for this year.

The spatio-temporal model without temperature forecasts an exponential decay towards the average COG during the forecast period (i.e., after the last year of fitted data). The average COG for *G. chalcogrammus* has shifted northward over 100 km from 1982 to 2015, so the later forecasts decay towards an average level that is more poleward than fits to earlier years (e.g., the purple forecast

**TABLE 2** Estimated parameters from the spatio-temporal forecast model using temperature fitted to all available data (1982–2015), showing the first-order autoregressive coefficient for spatio-temporal variation ( $\rho_n$  and  $\rho_w$ ), the standard deviation of spatial variation ( $\sigma_{\omega n}$  and  $\sigma_{\omega w}$ ) and spatio-temporal variation ( $\sigma_{en}$  and  $\sigma_{ew}$ ), and the standard deviation for random-walk variation in annual intercepts ( $\sigma_n$  and  $\sigma_w$ ) for both group density (n) and average biomass for each encountered group (w)

Species	Numbers density $n(s, t)$				Average weight $w(s, t)$			
	$\rho_n$	$\sigma_{\omega n}$	$\sigma_{en}$	$\sigma_n$	$\rho_w$	$\sigma_{\omega w}$	$\sigma_{ew}$	$\sigma_w$
<i>Gadus chalcogrammus</i>	0.91	0.61	0.24	0.20	0.51	1.06	1.00	0.08
<i>G. macrocephalus</i>	0.55	0.64	0.45	0.00	0.36	0.49	0.70	0.13
<i>Hippoglossoides elassodon</i>	0.98	1.36	0.27	0.00	0.91	0.46	0.34	0.00
<i>Chionoecetes opilio</i>	0.85	1.59	0.41	0.07	0.71	1.07	0.64	0.05
<i>H. stenolepis</i>	0.58	0.66	0.39	0.09	0.49	0.31	0.33	0.03
<i>Limanda aspera</i>	1.00	2.29	0.33	0.00	0.77	0.93	0.36	0.00
<i>Pleuronectes quadrifasciatus</i>	0.98	1.55	0.30	0.00	0.19	0.75	0.00	0.09
<i>C. bairdi</i>	0.88	1.28	0.45	0.00	0.81	0.69	0.43	0.12
<i>Podothecus accipenserinus</i>	0.57	1.40	0.49	0.12	0.73	0.86	0.40	0.11
<i>Atherestes stomias</i>	0.81	1.96	0.45	0.07	0.68	0.81	0.47	0.23
<i>Hyas coarctatus</i>	0.90	1.20	0.42	0.00	0.79	1.11	0.34	0.21
<i>Myoxocephalus polyacanthocephalus</i>	0.97	0.83	0.41	0.02	0.93	0.00	0.25	0.10
<i>Lycodes palearis</i>	0.98	2.33	0.35	0.00	0.67	0.56	0.30	0.12
<i>M. jaok</i>	0.51	1.69	0.66	0.00	0.51	0.47	0.37	0.10
<i>Mallotus villosus</i>	0.22	1.26	1.08	0.11	0.29	0.38	0.44	0.13
<i>H. lyratus</i>	0.86	1.34	0.39	0.12	0.89	0.45	0.25	0.25
<i>Clupea pallasi</i>	0.44	0.88	0.99	0.00	0.34	0.51	0.75	0.00
<i>Pandalus eous</i>	0.60	1.67	0.54	0.18	0.80	0.86	0.44	0.27
<i>Paralithodes camtschaticus</i>	0.93	2.28	0.38	0.00	0.75	0.84	0.24	0.05
<i>L. brevipes</i>	0.73	2.51	0.55	0.00	0.97	0.93	0.35	0.00



**FIGURE 2** Forecasts of northward centre of gravity for *Gadus chalcogrammus* (solid line: forecasted value; shaded area:  $\pm 1$  SE) using a habitat envelope (top-left panel), VAST model without temperature (top-right panel), VAST model with temperature (bottom-left panel) or annual-regression estimator (bottom-right panel), fitted to data through 1995, 1999, 2003, 2007, 2011, 2015 (see bottom-right panel for colour key), and compared with a abundance-and-area weighted average (AAWA) estimator of the centroid based on data for all years (dashed black line) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

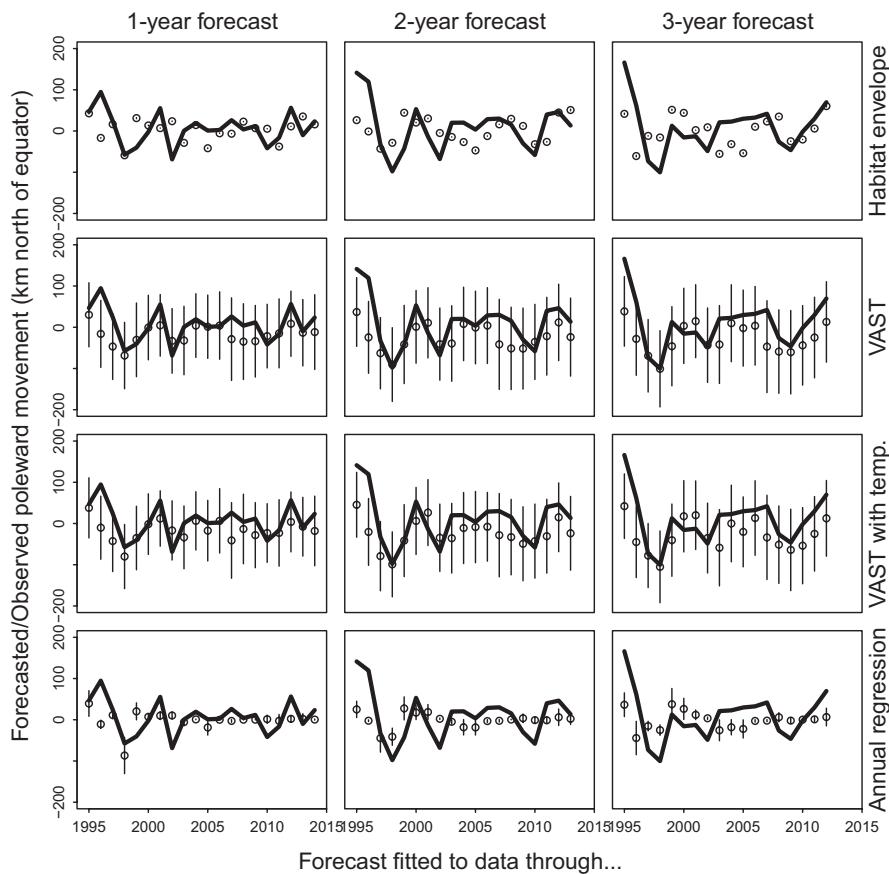
lines in 2015 are more northward than the blue forecast lines; Figure 2). By contrast, the spatio-temporal model with temperature generates forecasts that combine elements of both the VAST model without temperature and the habitat-envelope model. Specifically, the VAST model with temperature generates forecasts that follow an exponential decay towards the average COG of the historical period, but also evidence of interannual variability predicted by temperature patterns in each year (e.g., the purple forecast lines for the VAST model is less variable than the purple lines for the VAST model with temperature). For example, when fitting to data from 1982 to 1995, the COG forecast in 1997–1998 is higher than the forecast for 1999–2000, and this pattern is also seen in the habitat-envelope model. However, differences in forecast between VAST with and without temperature are small relative to the difference between VAST and the habitat-envelope model.

I next evaluate predictive skill when forecasting poleward movement 1, 2 or 3 years after the last year of fitted data for Alaska pollock. I first compare forecasted poleward movement using each estimator against the observed poleward movement using the abundance-and-area weighted estimator (Figure 3). This comparison shows that all models predict a poleward movement 1, 2 and 3 years forward using data 1982–1995, and this prediction is consistent with subsequent observations. However, the forecast interval is much too narrow for both the habitat-envelope and annual-regression estimators, and is more appropriate for the VAST forecasts with or without temperature. The VAST forecast also appears to be more accurate

than either habitat-envelope or annual-regression forecasts. For example, habitat envelope and annual regression predict a poleward shift when forecasting 1 year forward and fitting to data through 1999, whereas both VAST models forecast a southward shift and the southward shift is in fact observed (Figure 3). The improved forecast accuracy is clearly shown when calculating the correlation between forecasted and observed shifts, which is  $>0.5$  for 1-, 2- or 3-year forecasts using VAST with or without temperature, but is lower for habitat-envelope ( $<0.28$ ) or annual-regression ( $<0.39$ ) forecasts (Figure 4).

### 3.3 | Comparison of forecast skill across all 20 species

The correlation between forecasted and observed poleward movement is much greater for VAST models (either with or without temperature) than either habitat-envelope or annual-regression estimates across all 20 species, whether using 1-, 2- or 3-year forecasts (Figure 5). The variance explained by the habitat-envelope estimator for the median species is actually negative (Table 3), indicating that this forecast method has lower skill than a persistence forecast. Variance explained by the HE model is highly negative for species where it estimates a strong temperature response (e.g., see Figure 1: yellowfin sole (*Limanda aspera*, *Pleuronectidae*), shortfin eelpout (*Lycodes brevipes*, *Zoarcidae*) and wattled eelpout (*Lycodes palearis*, *Zoarcidae*)), but where the forecasted variance in distribution has



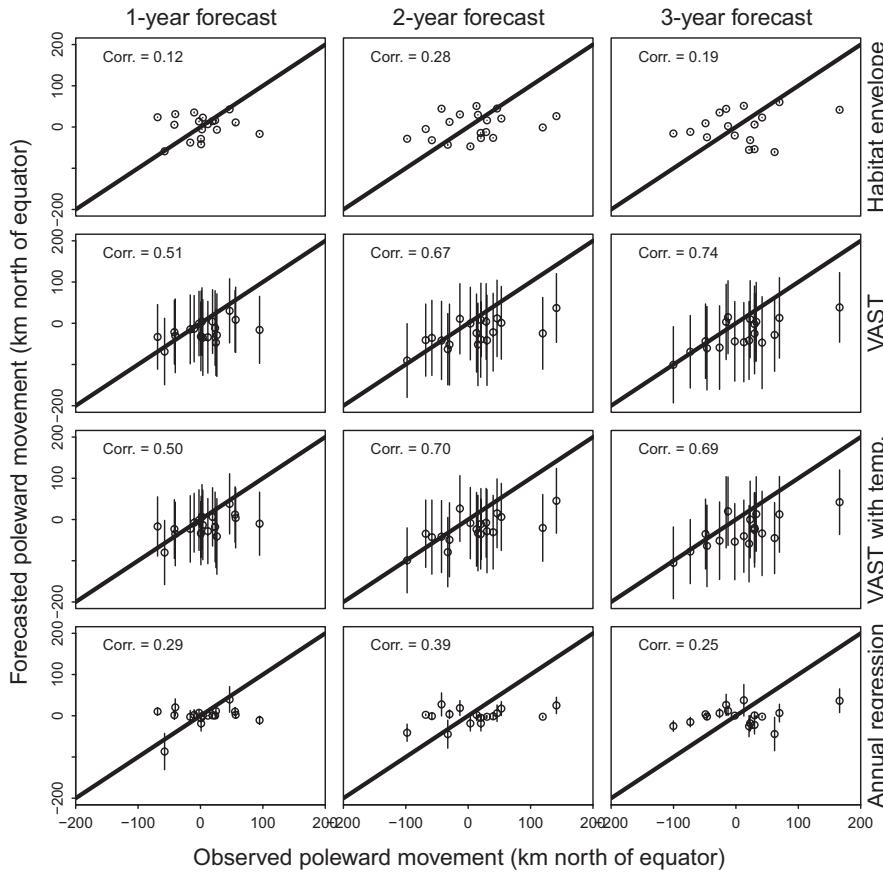
**FIGURE 3** Predicted vs. observed poleward movement (y-axis) when forecasting based on data through a given year (x-axis) for *Gadus chalcogrammus* using habitat envelope (top row), VAST model without temperature (2nd row), VAST model with temperature (3rd row), or annual-regression (4th row) forecast models and forecast 1 year (left column), 2 years (middle column), or 3 years (right column) ahead from the last year with available data. Each panel shows predicted poleward movement (circles),  $\pm 1$  SE (vertical lines centred at each circle) and observed poleward movement (black lines) using the abundance-and-area weighted estimator

little correlation with observed distribution shifts (Supporting Information Figures S1 and S2). The median variance explained is low but positive for the annual-regression estimator (0.02–0.06) and higher for both VAST models (0.08–0.25), and for these models the variance explained is higher for 3-year forecasts (0.20–0.25) relative to 2-year (0.14–0.16) or 1-year forecasts (0.08). Forecast models generally explain a larger portion of variance for 3-year forecasts because the error of the persistence forecast increases faster than the error of VAST forecasts (Supporting Information Appendix S3: Figure S3). The median variance explained is generally <25%, although individual species have higher variance explained (e.g., 69% for 3-year forecasts of Pacific cod (*Gadus macrocephalus*, Gadidae)), and VAST and annual-regression forecasts have greater skill than the persistence forecast. VAST has very similar performance with or without temperature, likely because the inclusion of a quadratic temperature response reduces the standard deviation of residual spatio-temporal variation by only 4%–6% on average (Supporting Information Appendix S3: Table S1), and predictions of change in COG are very similar between these two models (Supporting Information Figure S1). The habitat-envelope and annual-regression estimators perform poorly in terms of forecast interval coverage

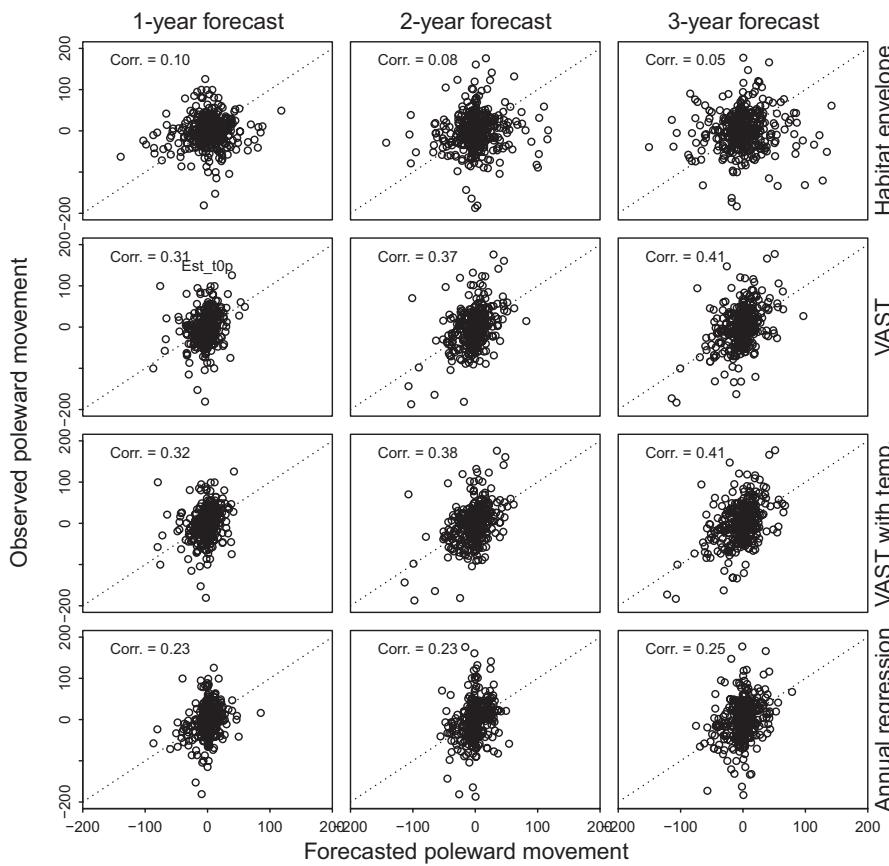
(Figure 6), where the observed poleward shift falls outside the 90% forecast interval for the majority of 1-, 2-, or 3-year forecasts. By contrast, the VAST models generate forecast intervals that contain the true value an appropriate proportion of the time, indicating that forecast intervals for this model are a useful measure of predictive uncertainty.

## 4 | DISCUSSION

Retrospective skill testing involves fitting ecological models to historical data, forecasting future changes and comparing forecasts with subsequent observations. I have demonstrated the potential role of retrospective skill testing for models forecasting distribution shifts by comparing two conventional estimators (habitat envelope, HE, and annual regression, AR) with a recently developed vector-autoregressive spatio-temporal (VAST) estimator. Using real-world data from 20 marine species in the Eastern Bering Sea and comparing performance with 400 retrospective analyses (20 forecast years for each species), I have showed that the conventional HE estimator has lower forecast skill than a simple “persistence” estimator. By



**FIGURE 4** Predicted (y-axis) vs. observed (x-axis) poleward movement for each comparison for *Gadus chalcogrammus* shown in Figure 3 (see Figure 3 caption for details). Each panel also shows the correlation between the predicted poleward movement and the observed poleward movement in the top-left corner and the one-to-one line (solid back). A perfect forecast model would have all pairs of predicted and observed shifts on the one-to-one line



**FIGURE 5** Predicted (x-axis) vs. observed (y-axis) poleward movement for all retrospective analyses for each of 20 species in the Eastern Bering Sea (see Figure 3 caption for details). Each panel also shows the correlation between predicted and observed shifts (top-left)

**TABLE 3** Variance explained relative to a persistence forecast,  $R^2(\Delta t)$ , for the median species (top row) or each individual species, for each forecast model (hybrid indicates the VAST model with temperature) for 1-, 2- or 3-year forecast intervals. A persistence forecast will have  $R^2(\Delta t) = 0.00$ , while a perfect forecast will have  $R^2(\Delta t) = 1.00$  (see Equation 15 Section 4 for details)

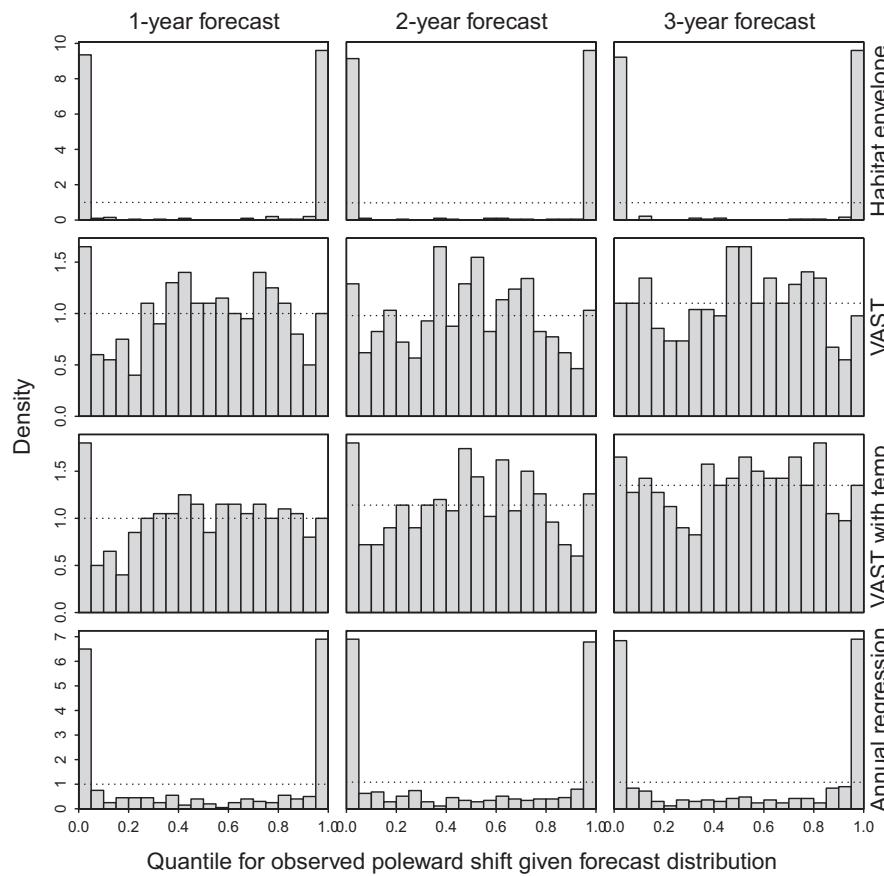
	1-year forecast				2-year forecast				3-year forecast			
	HE	VAST	Hybrid	AR	HE	VAST	Hybrid	AR	HE	VAST	Hybrid	AR
Median	-0.30	0.08	0.08	0.02	-0.42	0.16	0.14	0.06	-0.43	0.25	0.20	0.06
<i>Gadus chalcogrammus</i>	-0.29	0.25	0.23	0.00	0.02	0.43	0.47	0.15	-0.15	0.53	0.47	0.05
<i>G. macrocephalus</i>	-0.07	0.33	0.22	0.07	-0.13	0.62	0.51	0.16	-0.29	0.69	0.63	0.07
<i>Hippoglossoides elassodon</i>	-0.44	0.02	-0.01	0.07	-0.15	0.05	0.05	0.08	-0.64	0.07	-0.02	-0.09
<i>Chionoecetes opilio</i>	-0.51	0.05	0.07	-0.04	-0.4	0.13	0.16	0.08	-0.38	0.25	0.28	0.06
<i>H. stenolepis</i>	-0.70	0.30	0.19	0.17	-0.47	0.33	0.29	0.14	-0.48	0.31	0.36	0.21
<i>Limanda aspera</i>	-5.17	-0.03	-0.08	-0.31	-2.14	0.06	0.08	-0.20	-2.95	0.01	0.15	-0.19
<i>Pleuronectes quadrituberculatus</i>	-0.10	0.02	0.01	0.02	0.05	0.03	0.09	0.02	-0.14	0.02	0.02	0.05
<i>C. bairdi</i>	-0.72	0.07	0.09	-0.26	-0.62	0.15	0.16	-0.06	-0.55	0.22	0.23	0.01
<i>Podothecus accipenserinus</i>	-0.12	0.10	0.09	-0.14	-0.44	0.29	0.00	-0.06	-0.53	0.32	0.16	0.05
<i>Atheresthes stomias</i>	-0.17	0.16	0.5	0.53	-0.48	0.32	0.74	0.73	-0.32	0.43	0.77	0.73
<i>Hyas coarctatus</i>	0.16	0.09	0.09	0.18	0.02	0.15	0.13	0.06	0.1	0.25	0.21	0.13
<i>Myoxocephalus polyacanthocephalus</i>	-0.15	0.04	-0.04	-0.36	-0.8	0.03	0.06	-0.02	-0.65	0.11	0.16	-0.15
<i>Lycodes palearis</i>	-1.09	0.10	0.04	0.00	-1.4	0.17	0.06	0.06	-1.10	0.29	0.19	0.10
<i>M. jaok</i>	-0.82	0.21	0.11	0.09	-0.63	0.11	0.00	0.13	-0.65	0.27	0.20	0.13
<i>Mallotus villosus</i>	-0.06	-0.02	-0.01	0.12	-0.01	0.46	0.48	-0.09	0.05	0.19	0.28	0.31
<i>H. lyratus</i>	-0.34	-0.02	-0.06	0.00	-0.33	-0.01	-0.05	0.13	-0.35	0.16	0.11	0.10
<i>Clupea pallasi</i>	0.01	0.21	0.21	0.01	0.02	0.25	0.25	0.02	0.06	0.31	0.32	0.02
<i>Pandalus eous</i>	-0.31	0.01	0.09	0.20	-0.96	0.23	0.24	-0.07	-1.25	0.08	0.19	0.03
<i>Paralithodes camtschaticus</i>	0.20	0.01	-0.06	0.45	0.21	0.04	-0.02	0.41	0.24	0.04	0.02	0.40
<i>L. brevipes</i>	-3.54	0.13	0.04	-0.35	-2.76	0.19	0.15	-0.19	-3.15	0.25	0.18	-0.29

contrast, the annual-regression estimator generally explains a small (2%–6%) portion of variance in poleward movement, and the spatio-temporal estimators explain several times more variance (8%–25%). The spatio-temporal estimator is also the only method to estimate forecast intervals with appropriate width. I therefore conclude that the spatio-temporal estimator has suitable forecast skill, both in terms of accuracy and estimated uncertainty.

The VAST model performs better than the HE model for two main reasons. First, it accounts for temporally autocorrelated residuals in species distribution via the estimated autocorrelation parameters ( $\rho_n$  and  $\rho_w$ ). The importance of accounting for autocorrelated residuals when fitting species distribution models (Bahn & McGill, 2007; Dormann et al., 2007) or when forecasting population dynamics (Ives, Dennis, Cottingham, & Carpenter, 2003; Johnson et al., 2016) has been demonstrated previously, but has not been widely used when forecasting species distribution models. Secondly, the spatio-temporal model estimates the temporal variance of local changes in species density during the modelled period and uses this estimated variance to generate reasonable forecast intervals for future changes. As the explanatory power of covariates increases,

the variance of residual spatial and spatio-temporal variation will decrease and the performance of the HE and VAST models will converge. However, the variance of residual spatial and spatio-temporal variation remains large for all 20 species after accounting for a quadratic effect of bottom temperature (Supporting Information Appendix S3: Table S1). I therefore recommend using a spatio-temporal model as a testing ground for determining the relative explanatory power of covariates and residual environmental drivers, and recommend future research exploring additional covariates that could better explain historical variation in density.

Regarding the Eastern Bering Sea case-study presented here, results are consistent with prior analyses of this system in several important ways. First off, results show substantial variability in northward trends in COG for demersal species in the Eastern Bering Sea (see Supporting Information Figure S1), including instances where species are moving southward, or where a northward shift in distribution either is or is not predicted by a positive response to increasing regional temperatures (Mueter & Litzow, 2008) or decreasing cold pool area (Kotwicki & Lauth, 2013). Unlike previous analyses, however, the VAST model explored here estimates



**FIGURE 6** Quantile (see Equation 16 and its description for details) for the observed poleward movement (calculated using the abundance-and-area weighted estimator) relative to the predictive interval using habitat envelope, VAST models without or with temperature, and annual-regression estimators when forecasting either 1 (left column), 2 (middle column) or 3 (right column) years forward from the last year of fitted data, aggregated across results for all 20 species analysed in the Eastern Bering Sea. A model with appropriate forecast interval width will have a quantile distribution that is uniform (indicated by the dotted line in each panel), for example, such that a 50% forecast interval contains the observed poleward shift approximately 50% of the time, whereas a model with forecast intervals that are too small will have a quantile distribution concentrated near  $Q(\Delta t|t_{\text{final}}) = 0.0$  or  $Q(\Delta t|t_{\text{final}}) = 1.0$ .

autocorrelated variability in spatio-temporal patterns in population density ( $\rho_n$  and  $\rho_w$  in Table 2). This underlies autocorrelated variability in COG, and the latter can complicate interpretation of the correlation between COG and environmental covariates either when restricting analysis to a small subset of years (e.g., 2004–2009 in Hollowed et al. 2012) or when testing significance in a linear regression (e.g., Mueter & Litzow, 2008). Finally, this study did not explore several mechanisms that have invoked to explain distribution shifts in this region, for example, density-dependent habitat selection (Spencer, 2008; Thorson, Rindorf, Gao, Hanselman, & Winker, 2016) or age-specific temperature responses (Barbeaux & Hollowed, 2018; Thorson, Ianelli, et al., 2017). I therefore recommend future work exploring whether density dependence or age-structure can improve forecast skill in this or other regions.

Environmental management typically requires predictions over seasonal (<1 year), short-term (1–5 year), medium-term (5–50 year) and long-term (>50 year) planning horizons, and models may be more or less skilful at forecasts for each time-scale. Seasonal forecasts of spatial distribution for fished species can be useful for fishers to modify fleet behaviour, timing, and fishing techniques, and have been useful for prawn aquaculture in northeast Queensland, tuna fishing the Great Australia Bight, and lobster fishing the Gulf of Maine (see Tommasi et al. 2017 for a review). Predictive skill for seasonal forecasts has been extensively tested, for example, showing improved predictions of sardine distribution in the California Current relative to a persistence forecast (Kaplan et al., 2016). Although there are

sufficient data to conduct retrospective skill testing over seasonal or short-term forecasts, there is surprisingly little research using skill testing to compare performance among alternative models. I therefore recommend skill testing to either identify which models to use for different planning horizons, or how to weight predictions from multiple models when using an ensemble model for forecasting (e.g., Anderson et al., 2017).

Finally, this study highlights the need for increased skill testing for models used for short- and long-term forecasting of distribution shifts. For example, the decreased performance of a habitat-envelope model relative to a persistence forecast will likely surprise many readers. However, the poor performance of a habitat-envelope model is in-line with recent research showing that historical distribution shifts are poorly explained by local or regional temperature for Alaska pollock (Thorson, Ianelli, et al., 2017) and other groundfish in the Eastern Bering Sea (Kotwicki & Lauth, 2013). I recommend using retrospective analyses as a testing ground to improve probabilistic methods for forecasting short-term distribution shifts, and in particular exploring multispecies models for spatio-temporal dynamics (Latimer, Banerjee, Sang, Mosher, & Silander, 2009; Thorson & Barnett, 2017; Thorson, Rindorf, et al. 2016; Warton et al., 2015). In cases with strong species interactions or similar responses to shared environmental drivers, the information available in multispecies data sets may improve predictability for future dynamics (Ovaskainen et al., 2017; Schliep et al., 2018; Thorson, Munch, & Swain, 2017; Trainor et al., 2014). I therefore hypothesize that multispecies

spatio-temporal models will improve short-term forecasts of the northward distribution shift for commercial fishes in this region relative to the single-species forecast models explored here. Finally, I recommend further research using forecasting models to predict distribution based on future emission scenarios, for example, for use in future climate forecasts (Hollowed et al., 2009), but suggests that these methods should be skill-tested prior to use for this task.

## ACKNOWLEDGEMENTS

I thank K. Kristensen for developing the Template Model Builder software, without which the vector-autoregressive spatio-temporal model would not be feasible to fit. I also thank the many scientists who collected data in the Eastern Bering Sea shelf bottom trawl survey used here, which is an invaluable resource for exploring distribution shifts in marine species. Finally, I thank S. Kotwicki, J. Hastie, M. McClure and two anonymous reviewers for helpful comments on an earlier draft and grant #15-027 from the NOAA Habitat Assessment Improvement Plan that supported documentation for a precursor to package VAST.

## DATA ACCESSIBILITY

All data used are publicly available from the Alaska Fisheries Science Center at [http://www.afsc.noaa.gov/RACE/groundfish/survey\\_data/data.htm](http://www.afsc.noaa.gov/RACE/groundfish/survey_data/data.htm).

## ORCID

James T. Thorson  <http://orcid.org/0000-0001-7415-1010>

## REFERENCES

Adams, C. F., Alade, L. A., Legault, C. M., O'Brien, L., Palmer, M. C., Sosebee, K. A., & Traver, M. L. (2018). Relative importance of population size, fishing pressure and temperature on the spatial distribution of nine Northwest Atlantic groundfish stocks. *PLoS ONE*, 13, e0196583. <https://doi.org/10.1371/journal.pone.0196583>

Anderson, S. C., Cooper, A. B., Jensen, O. P., Minto, C., Thorson, J. T., Walsh, J. C., ... Selig, E. R. (2017). Improving estimates of population status and trend with superensemble models. *Fish and Fisheries*, 18, 732–741. <https://doi.org/10.1111/faf.12200>

Bahn, V., & McGill, B. J. (2007). Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, 16, 733–742. <https://doi.org/10.1111/j.1466-8238.2007.00331.x>

Barbeaux, S. J., & Hollowed, A. B. (2018). Ontogeny matters: Climate variability and effects on fish distribution in the eastern Bering Sea. *Fisheries Oceanography*, 27, 1–15. <https://doi.org/10.1111/fog.12229>

Cheung, W. W. L., Lam, V. W. Y., & Pauly, D. (2008). Dynamic bioclimate envelope model to predict climate-induced changes in distribution of marine fishes and invertebrates. In W. W. L. Cheung, V. W. Y. Lam & D. Pauly (Eds.), *Modelling present and climate-shifted distribution of marine fishes and invertebrates*, Vol. 16. Fisheries Centre Research Reports, 3rd edn (pp. 4–50). Vancouver, BC: Sea Around Us Project, Fisheries Centre, UBC.

Cheung, W. W. L., Lam, V. W. Y., Sarmiento, J. L., Kearney, K., Watson, R., & Pauly, D. (2009). Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries*, 10, 235–251. <https://doi.org/10.1111/j.1467-2979.2008.00315.x>

Cheung, W. W. L., Lam, V. W. Y., Sarmiento, J. L., Kearney, K., Watson, R., Zeller, D., & Pauly, D. (2010). Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. *Global Change Biology*, 16, 24–35. <https://doi.org/10.1111/j.1365-2486.2009.01995.x>

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., ... Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30, 609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>

Eskildsen, A., le Roux, P. C., Heikkinen, R. K., Høye, T. T., Kissling, W. D., Pöyry, J., ... Luoto, M. (2013). Testing species distribution models across space and time: High latitude butterflies and recent warming. *Global Ecology and Biogeography*, 22, 1293–1303. <https://doi.org/10.1111/geb.12078>

Eveson, J. P., Hobday, A. J., Hartog, J. R., Spillman, C. M., & Rough, K. M. (2015). Seasonal forecasting of tuna habitat in the Great Australian Bight. *Fisheries Research*, 170, 39–49. <https://doi.org/10.1016/j.fishres.2015.05.008>

Hare, J. A., Alexander, M. A., Fogarty, M. J., Williams, E. H., & Scott, J. D. (2010). Forecasting the dynamics of a coastal fishery species using a coupled climate–population model. *Ecological Applications*, 20, 452–464. <https://doi.org/10.1890/08-1863.1>

Hitch, A. T., & Leberg, P. L. (2007). Breeding distributions of North American bird species moving north as a result of climate change. *Conservation Biology*, 21, 534–539. <https://doi.org/10.1111/j.1523-1739.2006.00609.x>

Hollowed, A. B., Barbeaux, S. J., Cokelet, E. D., Farley, E., Kotwicki, S., Ressler, P. H., ... Wilson, C. D. (2012). Effects of climate variations on pelagic ocean habitats and their role in structuring forage fish distributions in the Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 65–70, 230–250. <https://doi.org/10.1016/j.dsr2.2012.02.008>

Hollowed, A. B., Bond, N. A., Wilderbuer, T. K., Stockhausen, W. T., A'mar, Z. T., Beamish, R. J., ... Schirripa, M. J. (2009). A framework for modelling fish and shellfish responses to future climate change. *ICES Journal of Marine Science*, 66, 1584–1594. <https://doi.org/10.1093/icesjms/fsp057>

Ives, A. R., Dennis, B., Cottingham, K. L., & Carpenter, S. R. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological Monographs*, 73, 301–330. [https://doi.org/10.1890/0012-9615\(2003\)073\[0301:ECSAEI\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2003)073[0301:ECSAEI]2.0.CO;2)

Johnson, K. F., Councill, E., Thorson, J. T., Brooks, E., Methot, R. D., & Punt, A. E. (2016). Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts? *Fisheries Research*, 183, 222–232. <https://doi.org/10.1016/j.fishres.2016.06.004>

Kai, M., Thorson, J. T., Piner, K. R., & Maunder, M. N. (2017). Spatiotemporal variation in size-structured populations using fishery data: An application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, 74, 1765–1780. <https://doi.org/10.1139/cjfas-2016-0327>

Kaplan, I. C., Williams, G. D., Bond, N. A., Hermann, A. J., & Siedlecki, S. A. (2016). Cloudy with a chance of sardines: Forecasting sardine distributions using regional climate models. *Fisheries Oceanography*, 25, 15–27. <https://doi.org/10.1111/fog.12131>

Kaschner, K., Watson, R., Trites, A. W., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, 316, 285–310. <https://doi.org/10.3354/meps316285>

Kass, R. E., & Steffey, D. (1989). Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes

models). *Journal of the American Statistical Association*, 84, 717–726. <https://doi.org/10.2307/2289653>

Kelly, A. E., & Goulden, M. L. (2008). Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11823–11826. <https://doi.org/10.1073/pnas.0802891105>

Kleisner, K. M., Fogarty, M. J., McGee, S., Hare, J. A., Moret, S., Perretti, C. T., & Saba, V. S. (2017). Marine species distribution shifts on the U.S. Northeast Continental Shelf under continued ocean warming. *Progress in Oceanography*, 153, 24–36. <https://doi.org/10.1016/j.pocean.2017.04.001>

Kotwicki, S., & 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 Research Part II: Topical Studies in Oceanography*, 94, 231–243. <https://doi.org/10.1016/j.dsr2.2013.03.017>

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and laplace approximation. *Journal of Statistical Software*, 70, 1–21. <https://doi.org/10.18637/jss.v070.i05>

Latimer, A. M., Banerjee, S., Sang, H. Jr, Mosher, E. S., & Silander, J. A. Jr (2009). Hierarchical models facilitate spatial analysis of large data sets: A case study on invasive plant species in the northeastern United States. *Ecology Letters*, 12, 144–154. <https://doi.org/10.1111/j.1461-0248.2008.01270.x>

Lauth, R. R., & Conner, J. (2016). *Results of the 2013 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate resources*. NMFS-AFSC-331, 155 pp. Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423–498. <https://doi.org/10.1111/j.1467-9868.2011.00777.x>

Lo, N. C., Jacobson, L. D., & Squire, J. L. (1992). Indices of relative abundance from fish spotter data based on delta-lognormal models. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 2515–2526. <https://doi.org/10.1139/f92-278>

Mellin, C., Lurgi, M., Matthews, S., MacNeil, M. A., Caley, M. J., Bax, N., ... Fordham, D. A. (2016). Forecasting marine invasions under climate change: Biotic interactions and demographic processes matter. *Biological Conservation*, 204, 459–467. <https://doi.org/10.1016/j.biocon.2016.11.008>

Merow, C., Latimer, A. M., Wilson, A. M., McMahon, S. M., Rebelo, A. G., & Silander, J. A. (2014). On using integral projection models to generate demographically driven predictions of species' distributions: Development and validation using sparse data. *Ecography*, 37, 1167–1183. <https://doi.org/10.1111/ecog.00839>

Mills, K. E., Pershing, A. J., Brown, C. J., Chen, Y., Chiang, F.-S., Holland, D., ... Wahle, R. (2013). Fisheries management in a changing climate: Lessons from the 2012 ocean heat wave in the Northwest Atlantic. *Oceanography*, 26, 191–195.

Mueter, F. J., & Litzow, M. A. (2008). Sea ice retreat alters the biogeography of the Bering Sea continental shelf. *Ecological Applications*, 18, 309–320. <https://doi.org/10.1890/07-0564.1>

Nye, J. A., Link, J. S., Hare, J. A., & Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Marine Ecology Progress Series*, 393, 111–129. <https://doi.org/10.3354/meps08220>

Olsen, E., Fay, G., Gaichas, S., Gamble, R., Lucey, S., & Link, J. S. (2016). Ecosystem model skill assessment. Yes we can! *PLoS ONE* 11, e0146467. <https://doi.org/10.1371/journal.pone.0146467>

Ovaskainen, O., Tikhonov, G., Dunson, D., Grøtan, V., Engen, S., Sæther, B.-E., & Abrego, N. (2017). How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings of the Royal Society B*, 284, 20170768. <https://doi.org/10.1098/rspb.2017.0768>

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, 37, 637–669. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110100>

Pinsky, M. L., & Fogarty, M. (2012). Lagged social-ecological responses to climate and range shifts in fisheries. *Climatic Change*, 115, 883–891. <https://doi.org/10.1007/s10584-012-0599-x>

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., & Levin, S. A. (2013). Marine taxa track local climate velocities. *Science*, 341, 1239–1242. <https://doi.org/10.1126/science.1239352>

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramírez-Albores, J. E., Bustamante, R. O., & Badano, E. I. (2016). Improved predictions of the geographic distribution of invasive plants using climatic niche models. *PLoS ONE*, 11, e0156029. <https://doi.org/10.1371/journal.pone.0156029>

Schliep, E. M., Lany, N. K., Zarnetske, P. L., Schaeffer, R. N., Orians, C. M., Orwig, D. A., & Preisser, E. L. (2018). Joint species distribution modelling for spatio-temporal occurrence and ordinal abundance data. *Global Ecology and Biogeography*, 27, 142–155. <https://doi.org/10.1111/geb.12666>

Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., ... Zimmermann, N. E. (2012). How to understand species' niches and range dynamics: A demographic research agenda for biogeography. *Journal of Biogeography*, 39, 2146–2162. <https://doi.org/10.1111/j.1365-2699.2012.02737.x>

Shelton, A. O., Thorson, J. T., Ward, E. J., & Feist, B. E. (2014). Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, 71, 1655–1666. <https://doi.org/10.1139/cjfas-2013-0508>

Spencer, P. D. (2008). Density-independent and density-dependent factors affecting temporal changes in spatial distributions of eastern Bering Sea flatfish. *Fisheries Oceanography*, 17, 396–410. <https://doi.org/10.1111/j.1365-2419.2008.00486.x>

Sugeno, M., & Munch, S. B. (2013). A semiparametric Bayesian approach to estimating maximum reproductive rates at low population sizes. *Ecological Applications*, 23, 699–709. <https://doi.org/10.1890/12-0453.1>

Swab, R. M., Regan, H. M., Matthies, D., Becker, U., & Bruun, H. H. (2015). The role of demography, intra-species variation, and species distribution models in species' projections under climate change. *Ecography*, 38, 221–230. <https://doi.org/10.1111/ecog.00585>

Theobald, E. J., Breckheimer, I., & HilleRisLambers, J. (2017). Climate drives phenological reassembly of a mountain wildflower meadow community. *Ecology*, 98, 2799–2812. <https://doi.org/10.1002/ect.1996>

Thorson, J. T. (2018). Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. *Canadian Journal of Fisheries and Aquatic Sciences*, 75, 1369–1382. <https://doi.org/10.1139/cjfas-2017-0266>

Thorson, J. T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. *Fisheries Research*, 210, 143–161. <https://doi.org/10.1016/j.fishres.2018.10.013>

Thorson, J. T., & Barnett, L. A. K. (2017). Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. *ICES Journal of Marine Science*, 74, 1311–1321. <https://doi.org/10.1093/icesjms/fsw193>

Thorson, J. T., Ianelli, J. N., & Kotwicki, S. (2017). The relative influence of temperature and size-structure on fish distribution shifts: A case-study on Walleye pollock in the Bering Sea. *Fish and Fisheries*, 18, 1073–1084. <https://doi.org/10.1111/faf.12225>

Thorson, J. T., Ianelli, J. N., Larsen, E. A., Ries, L., Scheuerell, M. D., Szuwalski, C., & Zipkin, E. F. (2016). Joint dynamic species distribution models: A tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, 25, 1144–1158. <https://doi.org/10.1111/geb.12464>

Thorson, J. T., Munch, S. B., & Swain, D. P. (2017). Estimating partial regulation in spatiotemporal models of community dynamics. *Ecology*, 98, 1277–1289. <https://doi.org/10.1002/ecy.1760>

Thorson, J. T., Ono, K., & Munch, S. B. (2014). A Bayesian approach to identifying and compensating for model misspecification in population models. *Ecology*, 95, 329–341. <https://doi.org/10.1890/13-0187.1>

Thorson, J. T., Pinsky, M. L., & Ward, E. J. (2016). Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. *Methods in Ecology and Evolution*, 7, 990–1002. <https://doi.org/10.1111/2041-210X.12567>

Thorson, J. T., Rindorf, A., Gao, J., Hanselman, D. H., & Winker, H. (2016). Density-dependent changes in effective area occupied for seabottom-associated marine fishes. *Proceedings of the Royal Society B*, 283, 20161853. <https://doi.org/10.1098/rspb.2016.1853>

Tommasi, D., Stock, C. A., Hobday, A. J., Methot, R., Kaplan, I. C., Eveson, J. P., ... Werner, F. E. (2017). Managing living marine resources in a dynamic environment: The role of seasonal to decadal climate forecasts. *Progress in Oceanography*, 152, 15–49. <https://doi.org/10.1016/j.pocean.2016.12.011>

Trainor, A. M., Schmitz, O. J., Ivan, J. S., & Shenk, T. M. (2014). Enhancing species distribution modeling by characterizing predator-prey interactions. *Ecological Applications*, 24, 204–216. <https://doi.org/10.1890/13-0336.1>

Tredennick, A. T., Hooten, M. B., Aldridge, C. L., Homer, C. G., Kleinhesselink, A. R., & Adler, P. B. (2016). Forecasting climate change impacts on plant populations over large spatial extents. *Ecosphere*, 7, e01525. <https://doi.org/10.1002/ecs2.1525>

Vitart, F. (2014). Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. <https://doi.org/10.1002/qj.2256>

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*, 30, 766–779. <https://doi.org/10.1016/j.tree.2015.09.007>

Zurell, D. (2017). Integrating demography, dispersal and interspecific interactions into bird distribution models. *Journal of Avian Biology*, 48, 1505–1516. <https://doi.org/10.1111/jav.01225>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Thorson JT. Forecast skill for predicting distribution shifts: A retrospective experiment for marine fishes in the Eastern Bering Sea. *Fish Fish*. 2019;20: 159–173. <https://doi.org/10.1111/faf.12330>