**TECHNICAL ADVANCE**

Global Change Biology WILEY

# Empirical orthogonal function regression: Linking population biology to spatial varying environmental conditions using climate projections

James T. Thorson[1] | Wei Cheng[2,3] | Albert J. Hermann[2,3] | James N. Ianelli[4] | Michael A. Litzow[5] | Cecilia A. O'Leary[6] | Grant G. Thompson[4]

[1]Habitat and Ecological Processes Research Program, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA

[2]Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA, USA

[3]Pacific Marine Environmental Laboratory, NOAA, Seattle, WA, USA

[4]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA

[5]College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Kodiak, AK, USA

[6]School of Aquatic and Fisheries Sciences, University of Washington, Seattle, WA, USA

**Correspondence**
James T. Thorson, Habitat and Ecological Processes Research Program, Alaska Fisheries Science Center, NMFS, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115, USA.
Email: James.Thorson@noaa.gov

## Abstract

Ecologists and oceanographers inform population and ecosystem management by identifying the physical drivers of ecological dynamics. However, different research communities use different analytical tools where, for example, physical oceanographers often apply rank-reduction techniques (a.k.a. empirical orthogonal functions [EOF]) to identify indicators that represent dominant modes of physical variability, whereas population ecologists use dynamical models that incorporate physical indicators as covariates. Simultaneously modeling physical and biological processes would have several benefits, including improved communication across sub-fields; more efficient use of limited data; and the ability to compare importance of physical and biological drivers for population dynamics. Here, we develop a new statistical technique, EOF regression, which jointly models population-scale dynamics and spatially distributed physical dynamics. EOF regression is fitted using maximum-likelihood techniques and applies a generalized EOF analysis to environmental measurements, estimates one or more time series representing modes of environmental variability, and simultaneously estimates the association of this time series with biological measurements. By doing so, it identifies a spatial map of environmental conditions that are best correlated with annual variability in the biological process. We demonstrate this method using a linear (Ricker) model for early-life survival ("recruitment") of three groundfish species in the eastern Bering Sea from 1982 to 2016, combined with measurements and end-of-century projections for bottom and sea surface temperature. Results suggest that (a) we can forecast biological dynamics while applying delta-correction and statistical downscaling to calibrate measurements and projected physical variables, (b) physical drivers are statistically significant for Pacific cod and walleye pollock recruitment, (c) separately analyzing physical and biological variables fails to identify the significant association for walleye pollock, and (d) cod and pollock will likely have reduced recruitment given forecasted temperatures over future decades.

## 1 | INTRODUCTION

Individual variation in animal growth, survival, and reproductive output is driven by differences in local habitat, where habitat differences arise from both bottom-up processes affecting resource availability and top-down control via predation and anti-predatory behaviors. Local habitat is, in turn, affected by oceanographic and atmospheric processes that exhibit complex spatial correlations across large scales (often called "teleconnections"). Mobile fishes, birds, and mammals will select optimal habitats based on available information, while population-scale demographic rates (e.g., adult survival rate and production of juveniles) represent the average rate for individuals in varied habitats that are distributed across space. Therefore, variation in population-scale rates is not typically driven by conditions at any single location, but instead correlates with variation in regional conditions that are associated with modes of physical variability that affect habitat at many different locations. In the Pacific Ocean, for example, the El Niño Southern Oscillation (ENSO) represents oceanographic and atmospheric teleconnections that correlate surface temperature and other ocean variables between spatially disparate sites. In this ecosystem, skipjack tuna track spatial shifts in the distribution of warm waters such that their spatial distribution is correlated with ENSO. However, population-scale demographic rates are buffered against temperature conditions occurring at any single site due to their ability to maintain suitable conditions via movement (Lehodey, Bertignac, Hampton, Lewis, & Picaut, 1997). In this and other cases, there is a need to associate spatially integrated biological rates with spatially distributed measurements of physical conditions when forecasting biological responses under changing physical conditions (Bindoff et al., 2019; Payne et al., 2017; Tommasi et al., 2017).

Ecologists and oceanographers have characterized dominant modes of environmental variability (and resulting teleconnections across geographically distant sites) using rank-reduction techniques including empirical orthogonal function (EOF) analysis (e.g., Hermann et al., 2019). EOF analysis takes as input the many spatially distributed measurements for a given oceanographic or atmospheric process (which we collectively call "physical processes"), each taken over time (typically over many years). EOF analysis reduces these measurements down to one or more time series as well as a map of physical conditions that are associated with a positive or negative value of each time series (Grimmer, 1963). EOF analysis has been used to define many well-known oceanographic processes, including the ENSO, the Pacific Decadal Oscillation (PDO), and many others (Kidson, 1975). Decades of research has involved extracting these indices and then including them as a covariate in a subsequent biological analysis (Mantua, Hare, Zhang, Wallace, & Francis, 1997; O'Leary, Miller, Thorson, & Nye, 2018).

For many marine fishes, variation in productivity and sustainable yield is driven in particular by large interannual variation in the production of juvenile fish, termed "recruitment" (Cushing, 1990; Hjort, 1926). As a consequence, understanding and predicting this variation has been a major goal of fisheries oceanography and stock assessment for over 100 years (Smith, 2007). The stock–recruit paradigm seeks to predict recruitment based on the biomass of spawning fish and resulting production of larvae, and there is widespread evidence that stock–recruit relationships occur (Myers, 2001). Although debates continue regarding specific methods for attribution (Szuwalski et al., 2019), it is clear that residual variability around this stock–recruit relationship is substantial (Thorson, Jensen, & Zipkin, 2014) and research is ongoing regarding statistical and mechanistic relationships to explain this residual variation after accounting for the stock–recruit relationship (Maunder & Thorson, 2019). However, most attempts to correlate stock–recruit residuals with environmental conditions have used pre-existing indices of ecosystem variability (e.g., PDO, North Atlantic Oscillation, etc.) rather than estimating the map of physical conditions that are most correlated with a specific biological process. There remains a pressing need to identify optimal sites to monitor as an index of environmental conditions, or alternatively to generate an environmental index that is appropriate for a given species and demographic process.

In this study, we develop a novel method that connects spatially distributed measurements of physical conditions with a time-series regression model that represents average biological rates at regional scales. Specifically, this method uses likelihood-based statistical methods to simultaneously estimate parameters for a physics component (a multivariate generalization of EOF analysis) and a biology component (a linear regression of a biological rate), to identify spatially correlated (and potentially nonlocal) physical conditions that are correlated with a given biological process. We demonstrate the method by fitting EOF analysis to bottom and surface temperature measurements in the eastern Bering Sea as well as projections of these same variables from a regional ocean model while fitting a linearized stock–recruit model to stock assessment records for three groundfishes. By combining these two processes, the approach estimates the map of physical conditions that have maximum correlation with above-average recruits-per-spawning biomass while also providing an estimate of the statistical significance of this relationship that propagates uncertainty from the physical component. In doing so, we demonstrate that the physics component can assimilate both historical ocean measurements and projected future conditions while intercalibrating the two data streams. This model-based intercalibration of historical and projected physical conditions allows us to forecast the stock–recruit relationship under end-of-century

climate conditions. We conclude with a description of other arenas in which it might be useful to jointly apply EOF analysis and biological time-series model.

## 2 | METHODS

We seek to identify the spatially distributed physical conditions that are associated with a spatially integrated biological process so that we can better estimate and forecast links between spatially varying physics and population-scale outcomes. We therefore jointly fit a statistical model to multiple datasets using two separate components for physics and biology. Assimilating both physical and biological data simultaneously allows the variance in the physics analysis to be propagated when forecasting biological responses (Niu et al., 2014). The physics component involves a statistical generalization of EOF analysis, whereas the biology component follows a conventional linear regression. Both components are fitted simultaneously using maximum-likelihood techniques as a mixed-effects model, and we distribute code to apply these methods in other circumstances as R package EOFR

(https://github.com/James-Thorson/EOFR). Future developments could incorporate nonlinear regression techniques, although we leave this as a topic for future research. We explain each component in detail in the following sections (see summary in Figure 1) while adapting conventions for mathematical notation from Edwards and Auger-Méthé (2019).

### 2.1 | Physics component

The physics component fits to measurements of multiple physical processes that are distributed across space and time. Specifically, we fit to measurements of $n_c$ physical processes, which are each indexed by $c$, and use the following notation to define spatial and temporal variation:

1. *Spatial variation*: We define $n_g$ grid cells that are indexed by $g$ and which discretize a spatial domain $\Omega$ (i.e., $g \in \{1,2,\ldots,n_g\}$), where each grid cell has location $\mathbf{s}_g$ (i.e., $\mathbf{s}_g \in \Omega$);
2. *Temporal variation*: We define $n_t$ time intervals between time $t_{min}$ and $t_{max}$ (i.e., $t \in \{t_{min}, t_{min}+1, \ldots, t_{max}\}$).
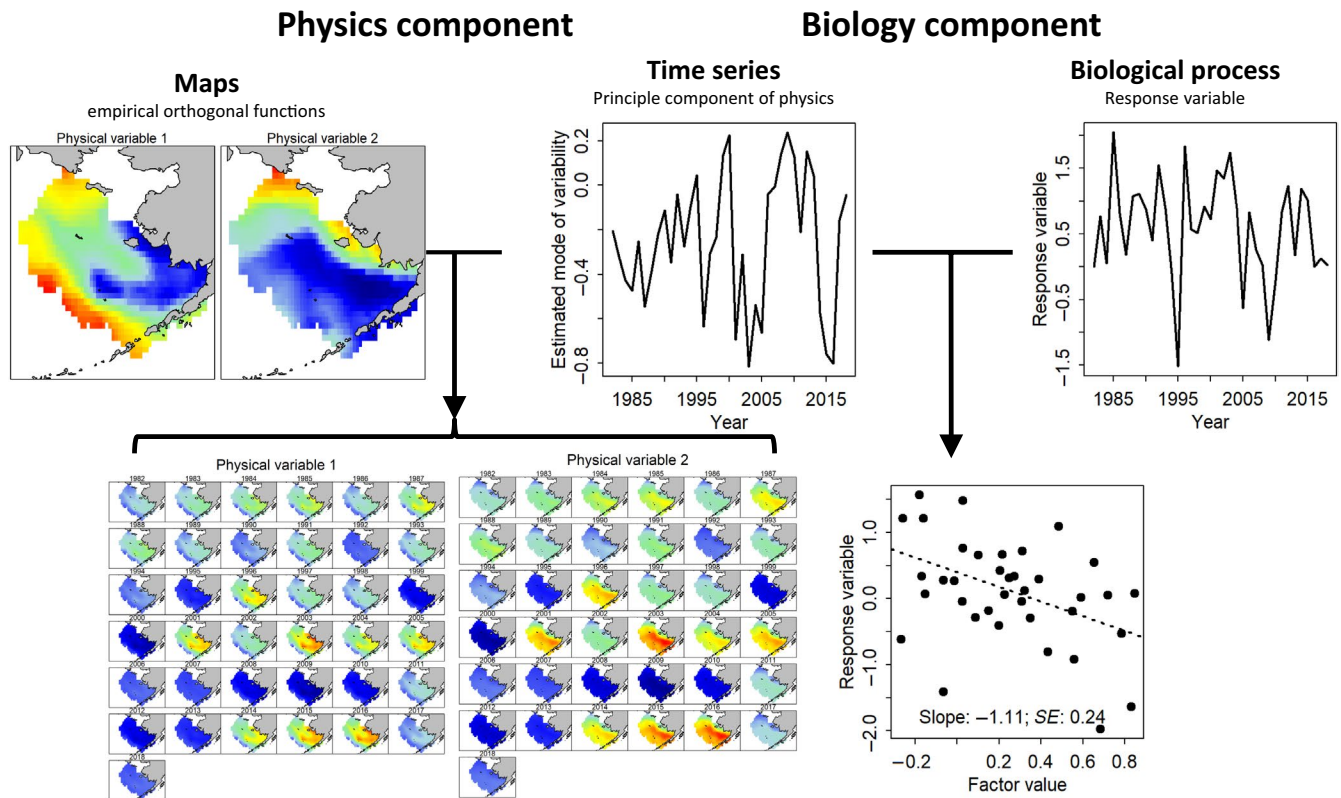


**FIGURE 1** Schematic showing linkage between the two components of the empirical orthogonal function (EOF) regression model: (a) the physics component (left-hand side), which includes one or more estimated time series representing the dominant mode of environmental variability (middle top panel) and an estimated spatial map (top-left panel) representing the response of each physical variable to that index, where the product of these is used to predict physical variables in each year (bottom left panels); and (b) the biology component (right-hand side), which includes the estimated time series as a predictor for a biological process measured over those same years (top-right panel). The linear model for the biological process (bottom-right panel) shows a statistically significant, negative relationship with the estimated time series; this indicates that when physical variables resemble their value in EOFs (top left panels), then this is associated with lower-than-average values for the biological process

We arrange measurements of physical processes in a matrix, where row $i$ includes the measurement $b_i$ as well as the category $c_i$, grid cell $g_i$, time $t_i$, and method $l_i$ that is used to obtain that measurement. We include index $l_i$ to allow us to fit to data from different measurement methods for a single physical variable. The physics component estimates spatial correlations and uses these to predict physical variables at unsampled locations. This allows the physics component to be fitted to physical variables that are spatially misaligned (i.e., different physical measurements can occur at different locations $g_i$) and unbalanced (i.e., measurements can be systematically missing in different locations $g_i$ and times $t_i$). The following development assumes that data are missing at random, but future studies could explore model-based approaches to relax that assumption (e.g., Conn, Thorson, & Johnson, 2017).

The physics component then fits a multivariate generalization of EOF analysis. This generalization assumes that spatial processes are on average more similar at geographically nearby than geographically distance locations (Tobler, 1970). Specifically, the model estimates one or more latent variables that include a spatial correlation function to approximate correlations among nearby locations (Banerjee, Gelfand, Finley, & Sang, 2008; Shelton, Thorson, Ward, & Feist, 2014). These spatial latent variables that are multiplied by a loadings matrix to approximate covariation among years, analogous to "factor" models in community ecology (Latimer, Banerjee, Sang, Mosher, & Silander, 2009; Thorson et al., 2015; Zuur, Fryer, Jolliffe, Dekker, & Beukema, 2003). This generalization has been explored previously (e.g., Thorson, Ciannelli, & Litzow, 2020), although these applications have not included a linkage between EOF indices and biological measurements.

This physics component estimates parameters for the loadings matrix $\mathbf{\Lambda}$ that contains the value $\lambda_{t,f}$ of index $f$ in each time $t$. It also estimates random effects $\varepsilon_{g,c,f}$ representing the expected response to index $f$ at location $g$ for each variable $c$. Any two locations $g_1$ and $g_2$ that have a large magnitude for response $\varepsilon_{g,c,f}$ for a given factor $f$ will tend to have correlated physical dynamics. When these two locations $g_1$ and $g_2$ are geographically distant, we refer to these correlations as teleconnections; these teleconnections can occur among locations for a single variable $c$, or among two variables $c_1$ and $c_2$, as well as beyond the decorrelation distance of the spatial correlation function that is used when estimating random effects. The physics component also estimates the variance of residual errors $\sigma_c^2$, the spatial decorrelation rate $\kappa$, and the intercept $\alpha_l$ and variance of spatial variation $\sigma_l^2$ for each type of measurement, where $\delta_{g,l}$ is a random effect representing spatial variation in average measurements for measurement type $l$ at each location $g$. Parameters are estimated via maximum likelihood (see Appendix A in Data S1 for more details regarding model structure and Appendix B in Data S1 for identifiability criteria). When there are two or more measurements $l$ for the same variable $c$, then the difference $\delta_{g,l_1} - \delta_{g,l_2}$ represents spatial variation in the expected difference between two measurements of the same variable; this is conceptually similar to statistical downscaling and the delta-change method (Gleick, 1986).

## 2.2 | Biological component

To demonstrate the linkage between physics and biology components as clearly as possible, we specify a very simple biological component. Specifically, we use a linear regression to represent biological dynamics; a linear relationship is commonly used to represent annual demographic rates as a function of annual environmental conditions. The linear model includes record $y_j$ for measurement $j$ of $n_j$ measurements, and minimizes the difference between these records and model predictions $y_j^*$:

$$y_j \sim \text{Normal}\left(y_j^*, \tau^2\right). \tag{1}$$

In the following, we use a normal distribution with variance $\tau^2$, although future developments could explore a generalized linear model with a link function and distribution from the exponential family, or a nonlinear model linking covariates and response variable. The linear model also includes $n_k$ covariates $x_{j,k}$ that represent measured processes that explain the biological process beyond what is explained by physical variables included in the physics component.

Unlike conventional linear regression, however, we include an estimated mode of physical variability within our linear regression:

$$y_j^* = \underbrace{\beta_0 + \sum_{k=1}^{n_k} \beta_k x_{j,k}}_{\text{Conventional linear model}} + \underbrace{\gamma \lambda_{t_j,1}}_{\substack{\text{Linkage to} \\ \text{physical submodel}}}. \tag{2}$$

This model estimates parameters representing the intercept $\beta_0$, covariate effects $\beta_k$, residual variance $\tau^2$, as well as the parameter $\gamma$ governing the cross-correlation between physics and biology components. When fixing $\gamma = 0$, the physics and biology components include no parameters in common and the model reduces to two independent models for biological and physical processes. When estimating $\gamma$, by contrast, the model can identify more terms in loadings matrix $\mathbf{\Lambda}$ (see Appendix B in Data S1 for details regarding identifiability criteria). Maximum-likelihood estimates for these additional coefficients represent the role of each index $\lambda_f$ in explaining residuals in $y_j$. By jointly estimating physics and biology components, the model can identify additional coefficients in $\lambda_f$ simultaneously with other biological covariates $\beta_k$. It therefore estimates a mode of physical variability (analogous to how the PDO or ENSO is conventionally estimated), except it does so in a way that customizes this variable to maximize its explanatory power for a given biological process.

## 2.3 | Case-study demonstration: Recruitment for eastern Bering Sea groundfishes

We demonstrate this joint model by fitting the widely used Ricker stock–recruit model to records of age-0 recruitment (in numbers) and spawning biomass in each year from 1982 to 2016 for three

commercially important groundfish species in the eastern Bering Sea, and then linking this model to end-of-century climate projections to forecast changes in recruitment for these species. The Ricker model represents density-dependent fecundity and early-life survival in fishes (see Appendix C in Data S1 for details), and it has been widely documented in fishes that compete for space and resources when spawning (Foss-Grant, Zipkin, Thorson, Jensen, & Fagan, 2016). We specifically link bottom and surface temperature to stock–recruit relationships for three groundfishes in the eastern Bering Sea: Pacific cod (*Gadus macrocephalus*), walleye pollock (*Gadus chalcogrammus*), and yellowfin sole (*Limanda aspera*). We note that these stock–recruit records are estimates from age-structured stock assessment models, with associated errors arising from sampling variability and model mis-specification. Treating estimates as data in this way results in several well-known biases (Brooks & Deroba, 2015; Dickey-Collas, Payne, Trenkel, & Nash, 2014; Ludwig & Walters, 1981), and we encourage future research that applies the EOF regression approach developed here simultaneously with estimating parameters for age-structured population models.

We analyze these stock and recruitment records jointly with surface and bottom temperatures. Specifically, we include bottom temperature measurements from net sensors attached to bottom trawl samples conducted by the Bering Sea shelf survey, conducted during late Spring and early summer from 1982 to 2018 (Lauth & Conner, 2014), and surface temperature records from the NOAA Extended Reconstructed Sea Surface Temperature (v.5) product (Huang et al., 2017), representing the January average from 1982 to 2018. We also include Regional Ocean Modeling System (ROMS) hindcasts of bottom and surface temperatures from 2009 to 2018 and ROMS projections of these variables every 10th year from 2020 to 2090. These ROMS hindcasts and projections are generated by a downscaled regional hindcast model (Bering10K-NPZ), which was itself driven by observed atmospheric and oceanic conditions from global reanalysis (the Climate Forecast System Reanalysis), and then projected to 2090 by downscaling a subset of CMIP5 Earth System models (Hermann et al., 2019; Kearney, Hermann, Cheng, Ortiz, & Aydin, 2020). We treat bottom and surface temperature as two measurements, and bottom and surface hindcasts/projections as two additional measurements ($n_I = 4$), and the model intercalibrates these multiple data streams using the overlap from 2009 to 2018 between measurements and hindcasts/projections. We estimate three modes of physical variability; future research could explore using model-selection techniques to determine the optimal number of modes, although we do not address the topic here.

Finally, we compare results using the new approach EOF regression with a conventional approach to linking spatially distributed measurements of physics as well as a time series representing a biological variable that integrates across the same spatial domain. To do so, we calculate the primary mode of variability for surface temperature records, and include this as a covariate in the same linearized Ricker model for stock–recruit records (Appendix D, Figure S1 in Data S1). We extract the statistical significance using this conventional approach, and compare it with the significance using EOF regression.

## 3 | RESULTS

Visualizing predictions of surface and bottom temperatures (Figure 2) based upon bottom-temperature measurements available for 1982–2018 and ROMS model hindcasts and projections available for 2009–2090 show that the physics component identifies interannual variation in the spatial extent of cold near-bottom waters (termed the "cold pool," and shown as blue areas in the first and third columns of Figure 2). In particular, 1999 had one of the largest measured cold pools while 2018 had one of the smallest cold pools on record. This pattern is captured by field measurements and ROMS simulation for both years, although field measurements in 1999 could also be extreme due to small changes in survey timing in that year. Predictions are slightly different between the measurements and ROMS predictions due to model-based intercalibration of these two data streams. Bottom temperatures in the inner domain and along the Aleutian Islands are then projected to substantially increase in 2040 and particularly by 2090. Similarly, surface temperatures are projected to greatly increase in the southern Bering Sea over time (Figure 2, second and fourth columns). The areas of greatest warming for bottom temperatures (near Nunivak Island and Bristol Bay) are geographically distant from areas with warming surface temperatures (near the Aleutian Islands), and this illustrates that EOF analysis can capture "teleconnections" arising among physical variables. When estimating modes of physical variability without a link to biological data (Appendix E in Data S1), the first mode represents an increase in bottom temperature in the inner domain (Appendix E, Figure S2b in Data S1), whereas the second and third modes collectively capture variation in the extent of the cold pool (Appendix E, Figure S2a in Data S1).

By contrast, the first factor varies by species when jointly fitting to physical variables and biological time series (Appendix E, Figure S3 in Data S1). This factor typically has a high value in 1999, which also has elevated recruits-per-spawning biomass for walleye pollock and Pacific cod (Figure S3 in Data S1, first two rows). The map associated with this first factor shows the variation in physical conditions that is most correlated with above-average recruits-per-spawning biomass for each species (Figure 3). This map illustrates the locations where unusual physical conditions are most likely to impact recruits-per-spawning biomass for each species, and it differs among species as expected given different biological characteristics of each species. Specifically, the model expects that Pacific cod and walleye pollock have elevated recruits-per-spawning biomass in years with cooler waters in the southern middle domain (as occurs in years with large cold-pool extent). Recruitment for both walleye pollock and Pacific cod is associated with cooler waters that extend nearshore from Bristol Bay to Nunivak Island, whereas Pacific cod recruitment is also associated with cooler waters offshore to St. George Island. Finally,
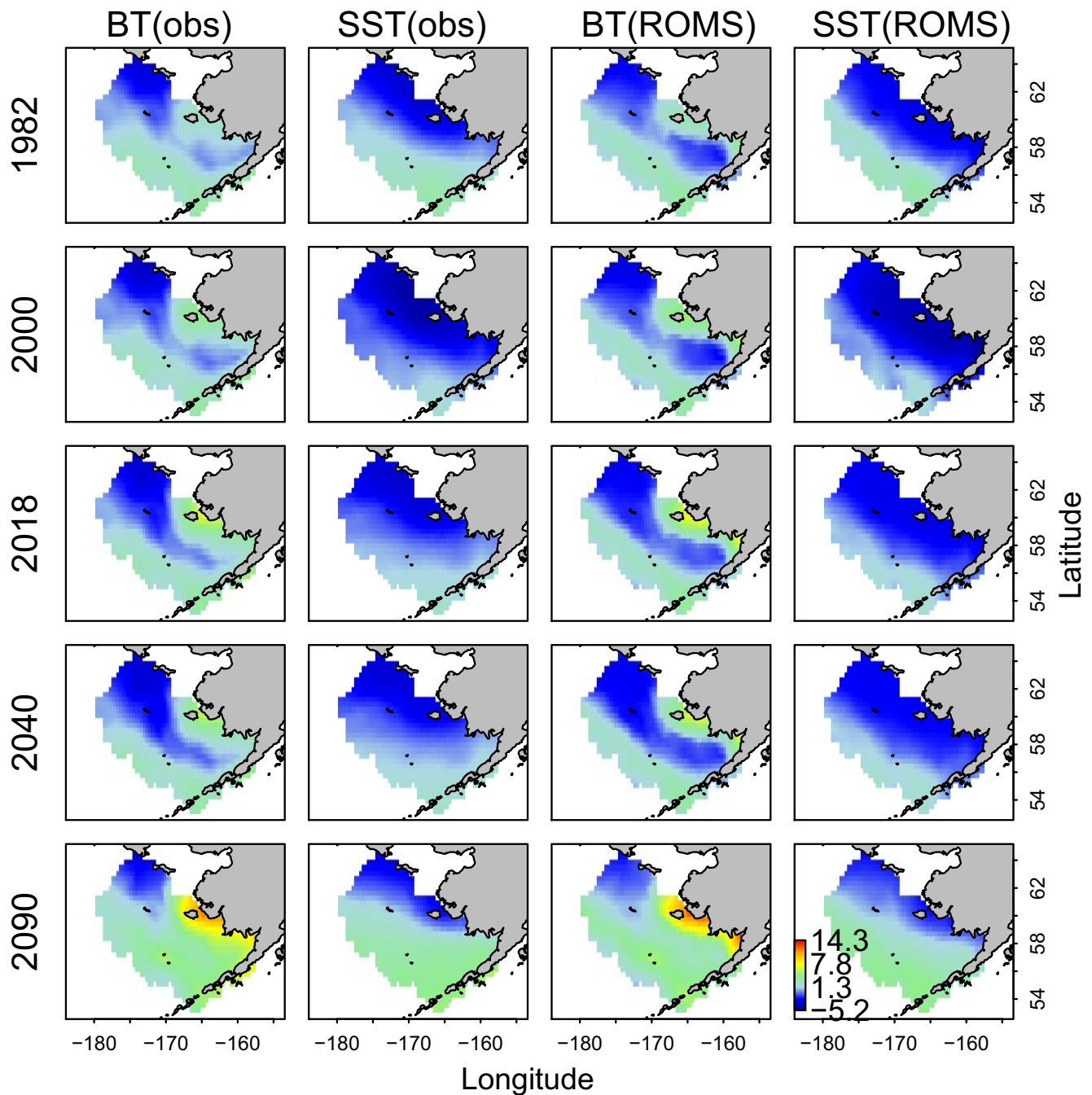
**FIGURE 2** Illustration of predicted bottom temperature (BT) and surface temperature (SST) using a model that is not linked to biological variation (i.e., $\gamma = 0$), showing predictions of both measurements and Regional Ocean Modeling System (ROMS) projections of each variable ($n_l = 4$, columns) in the eastern Bering Sea, fitted to measurements 1982–2016 as well as ROMS hindcasts and projections 2009–2018 and every 10th year 2020–2090. Here we specifically show evenly spaced years during historical and forecasted periods (1982/1999/2018 and 2040/2090). Differences among years (rows; e.g., from 1982 to 2090) for a given variable (column) visualize the regional expression of climate change; differences among predicted measurements for a given variable (e.g., first and third columns) show the model-based intercalibration of physical measurements and ROMS projections; differences among variables (e.g., first and second columns) show the varying expression of climate change for different physical variables

yellowfin sole recruitment is expected to be elevated under different environmental conditions, for example, elevated bottom temperatures inshore from Nunivak Island. These differences are also strongly apparent in surface-water associations, where walleye pollock and Pacific cod recruitment is expected to be elevated

given lower temperatures in the southern Bering Sea, while yellowfin sole is depressed during these same conditions. Elevated yellowfin recruitment is associated with opposite physical conditions to conditions with elevated walleye pollock and Pacific cod recruitment, and this inverse relationship arises in part due to the
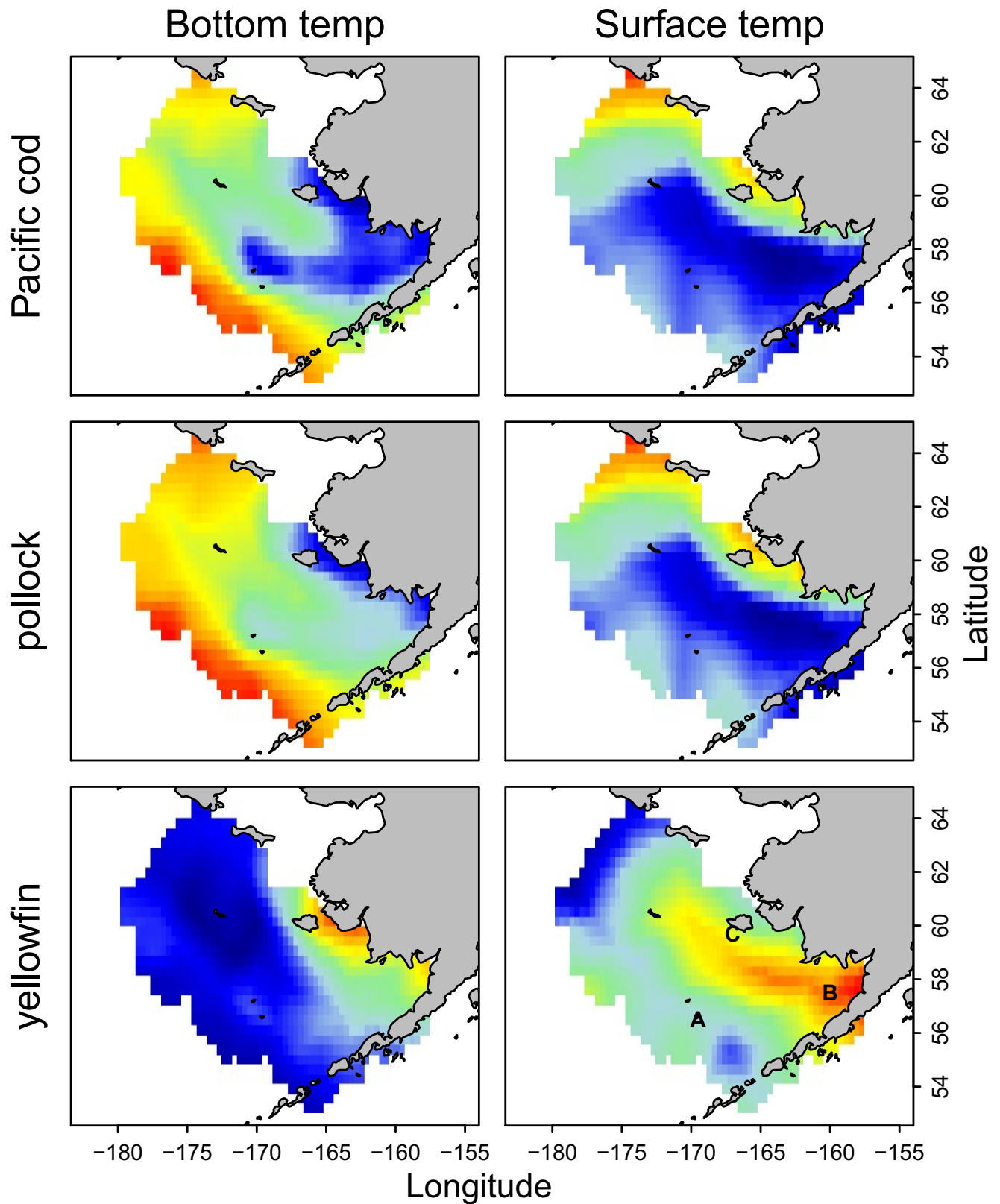
## Bottom temp

## Surface temp



**FIGURE 3** Visualization of estimated maps $\varepsilon_{g,c,1}$ for a given physical variable (column) for the factor that is simultaneously used as a covariate when fitting to stock–recruit data for Pacific cod, walleye pollock, or yellowfin sole (rows). Colors show locations that are positively (red) or negatively (blue) associated for a given physical variable with variation in log-recruits-per-spawning biomass for each species; the bottom right panel labels locations mentioned in text (A: St. George Island; B: Bristol Bay; C: Nunivak Island). We show plots for each species to highlight the different spatial locations that are associated with positive residuals in recruits-per-spawning biomass
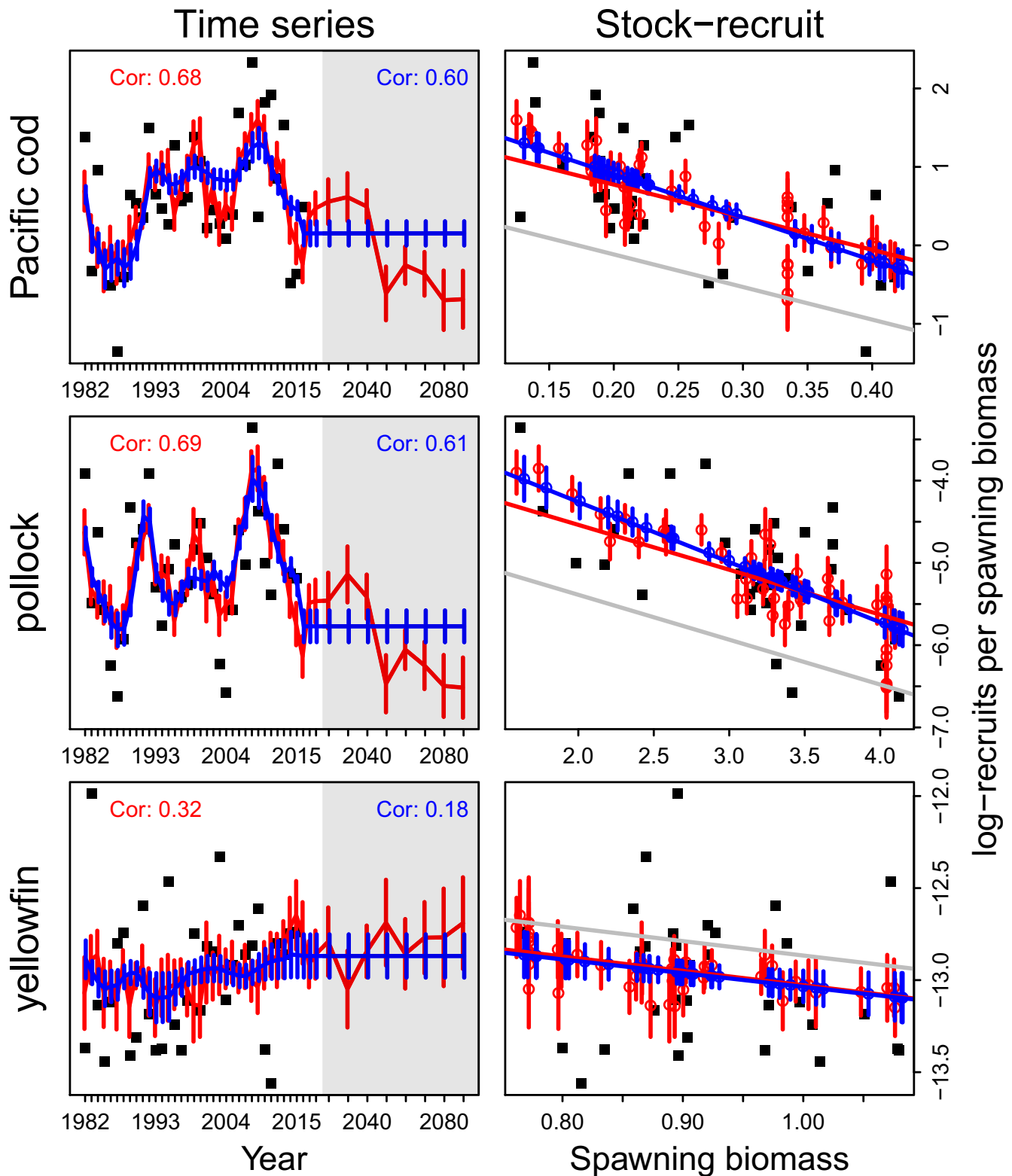
**FIGURE 4** Visualization of log-recruits-per-spawning biomass (*y*-axis) either as a function of year (left-column *x*-axis) or spawning biomass (right-column *x*-axis) for models fitted to stock–recruit data for each individual species (rows). For each species, we show the recorded value (black square), the estimate without cross-correlation, *γ* = 0 (blue line) or estimating cross-correlation *γ* (red line), where vertical lines represent ±1 *SE*. In each time series, we list the correlation between recorded values and predicted log-recruits-per-spawning biomass for models without (blue number) or with cross-correlation (red number). In each time-series panel, we indicate forecasted years (2020–2090) using gray shading. In each stock–recruit panel, we include a line showing the estimated stock–recruit relationship, which is linear given the parameterization of the Ricker model being used, and show the predicted stock–recruit relationship given estimated physical conditions in 2090 (gray line)

**TABLE 1** Parameter estimates for walleye pollock, Pacific cod, and yellowfin sole for models with and without a cross-correlation (columns), showing Akaike information criterion (AIC) and the number of parameters using maximum likelihood, and reporting estimated parameters (standard errors in parentheses) when using restricted maximum likelihood. For the linked slope parameter γ, we also report the *p* value for a two-sided Wald test after the standard error in parentheses

| | Pacific cod | | | Walleye pollock | | | Yellowfin sole | | |
|---|---|---|---|---|---|---|---|---|---|
| | Empirical orthogonal function (EOF) regression Cross-correlated | | | EOF regression Cross-correlated | | | EOF regression Cross-correlated | | |
| | Yes | No | Conventional | Yes | No | Conventional | Yes | No | Conventional |
| ΔAIC | 0 | 0.4 | | 0.4 | 0 | | 3.4 | 0 | |
| Number of parameters | 149 | 146 | | 149 | 146 | | 149 | 146 | |
| Ricker intercept, α | 1.91 (0.46) | 2.00 (0.34) | 1.85 (0.34) | −2.98 (0.63) | −2.80 (0.53) | −3.35 (0.56) | −12.39 (0.74) | −12.29 (0.66) | −12.20 (0.70) |
| Ricker slope, β | −4.15 (1.29) | −5.48 (1.27) | −4.70 (1.29) | −0.55 (0.18) | −0.73 (0.17) | −0.54 (0.18) | −0.78 (0.73) | −0.74 (0.72) | −0.85 (0.76) |
| Linked slope, γ | 0.93 (0.39), p = .02 | NA | −1.76 (0.93), p = .06 | 0.83 (0.38), p = .03 | NA | −2.10 (0.96), p = .03 | 0.24 (0.16), p = .13 | NA | 0.25 (0.55), p = .64 |
| Residual variation in physics, ln(σγ) | −0.14 (0.01) | −0.14 (0.01) | | −0.14 (0.01) | −0.14 (0.01) | | −0.14 (0.01) | −0.14 (0.01) | |
| Residual variation in biology, ln(σR) | −0.51 (0.13) | −0.43 (0.12) | | −0.55 (0.13) | −0.49 (0.13) | | −1.05 (0.13) | −1.02 (0.13) | |

restriction that environmental drivers of recruitment are associated with one of the leading modes of physical variability.

Finally, we compare estimates of early juvenile survival (log-recruits-per-spawning biomass) for a model with or without a linkage to regional physics (Figure 4). For all three species, including physical variables when predicting recruitment substantially improves the correlation between predicted recruitment and stock-assessment estimates of recruitment (from 0.60 to 0.68 for Pacific cod, from 0.61 to 0.69 for walleye pollock, and from 0.18 to 0.32 for yellowfin). Parameter γ representing this link is significant for pollock and cod (based on a two-sided Wald test), although the Akaike information criterion indicates that the model linking physics and biology is only parsimonious for Pacific cod (Table 1); the parameter is neither parsimonious nor significant for yellowfin sole. By contrast, physical drivers are only significant (using a two-sided Wald test) for walleye pollock and not significant for Pacific cod or yellowfin when using the conventional approach that first applies EOF analysis to surface temperatures and then uses the first mode of variability as a covariate in a stock–recruit model (see "Conventional" columns of Table 1).

The EOF regression model expected elevated Pacific cod and walleye pollock recruitment given cooler bottom temperatures in the southern middle domain, so it is unsurprising that recruitment for these two species is expected to decline over the coming decades. In particular, log-recruits-per-spawning biomass is expected to decline by 0.89 by 2090 for Pacific cod, corresponding to a nearly 59% decline in yield all else equal (i.e., yield-per-recruit, spawning biomass, recruitment density dependence remain constant), and by 0.85 for walleye Pollock, corresponding to a nearly 57% decline in yield all else equal. Alternatively, increased yellowfin recruits-per-spawning biomass is expected given elevated bottom and surface temperatures, and explains an increase in expected recruitment for this stock.

## 4 | DISCUSSION

In this paper, we explained the rationale for relating spatially integrated biological processes (e.g., stock–recruit relationships) to dominant modes of physical variability that are associated with atmospheric and oceanographic teleconnections. We then developed a statistical model that jointly fits to spatially distributed physical measurements and population-scale biological measurements. The resulting model estimates the map of physical conditions that are maximally correlated with a given biological process, and simultaneously performs delta-correction and statistical downscaling to intercalibrate field measurements and climate projections from earth systems models. Finally, we showed how this model can represent the widely used Ricker model for early-life survival in fishes, and used three case-study applications to show how the model can identify significant relationships between physics and juvenile survival. We confirmed that a conventional, separate analysis of physics and biological processes fails to identify one of the relationships as significant, and also highlight the different landscape of environmental

conditions that are favorable for recruitment of each species. We now discuss how EOF regression could be extended in future research, as well as why we hypothesize that it will help to identify stationary and skillful associations between physics and resulting biological outcomes.

Changing climate conditions present a challenge for environmental management (Karp et al., 2019). Climate is known to impact fish productivity and distribution (Brander, 2010; Finney et al., 2010; Holsman, Essington, Miller, Koen-Alonso, & Stockhausen, 2012; Lehodey et al., 2006), but it remains difficult to detect climate linkages that result in skillful, annual predictions of parameters that are used for fisheries management. We anticipate that EOF regression can be used to rapidly screen for physical conditions that are associated with a given biological process. Retrospective skill testing could then be used to identify which associations between physics and biology have been stable over recent years, and therefore are likely to persist over the next several years (Karp et al., 2019; Thorson, 2019). Some physics–biology associations may be skillful for short-term predictions; in other cases, estimated associations may be useful to target future research regarding biological mechanisms. Although we studied the stock–recruit relationship in this study, EOF regression could be extended to model a nonlinear relationship between physics and biology for a wide range of taxa; for example, informing optimal spatial planning or population viability analysis.

As climate change causes environmental conditions to exceed the climate envelope that prevailed during historical data collection, it becomes increasingly important to understand mechanisms governing the abundance, distribution, recruitment, productivity, and community interactions of natural populations. However, there are two major problems when linking physical variability to population processes: (a) environmental variables are often colinear such that individual effect of each variable is difficult to estimate (Dormann et al., 2013) and (b) it is difficult to attribute observed variability to causal mechanisms using short, noisy observational data. Using an example for fishes in the Bering Sea, water temperature and seasonal stratification patterns are correlated processes on the continental shelf (Stabeno et al., 2012). The former variable is expected to affect groundfish recruitment through thermal effects on physiological processes as reflected in growth and development rates (Laurel, Spencer, Iseri, & Copeman, 2016), whereas the latter is expected to affect recruitment by modulating the timing and magnitude of primary production (Sigler, Stabeno, Eisner, Napp, & Mueter, 2014). We speculate that the importance of these two processes will in some cases be identifiable by fitting to spatially distributed measurements of both temperature and stratification; identifying their relative importance will be facilitated whenever there is spatial variation that provides contrast to separate the impact of these colinear processes on a biological time series. Results when fitting to this expanded set of physical variables may, in turn, be useful for refining hypotheses for subsequent process studies for improved mechanistic understanding of climate impacts on population dynamics, or for relating changes in spatial distribution to regional climate variability.

We note several topics that warrant further research. These include the following:

1. Jointly fitting EOF regression with other modeling steps. For example, we could jointly fit an EOF to physics and an age-structured or integral-projection model (Merow et al., 2014). This would avoid using stock–recruit records as data as we did in this analysis.
2. Comparing EOF regression with approaches to analyzing physical modes of variability (see Appendix F in Data S1 for past approaches in the Bering Sea). For example, analysts have conventionally used oceanographic indices extracted from EOF analyses applied to physical data, and then included those in population models (see Appendix D in Data S1; Table 1). We recommend further performance comparisons of EOF regression with the approaches that apply EOF and biological analyses separately; our comparison for these three case studies suggests that EOF regression is a more efficient use of available data and can detect a significant relationship in cases that a sequential approach might otherwise miss.
3. Exploring implications for population management of case-study species. Each species analyzed here is commercially important and subject to regular assessment of sustainable catch limits and stock status. Therefore, incorporating physical information to predict juvenile survival could be useful for each stock, but would also require careful consideration to incorporate within existing management processes (see Appendix G in Data S1).

We hope to facilitate these and other research topics by making our code publicly available at https://github.com/James-Thorson/EOFR.

Another persistent problem in resolving environmental control of population variability in natural populations is the tendency for observed environment–biology correlations to arise and disappear over time (Myers, 1998). These "ephemeral" correlations are often interpreted as reflecting correlation that is incorrectly confused with causality (Deyle, May, Munch, & Sugihara, 2016). However, an alternative explanation is that of novel climates, that is, previously unobserved patterns of collinear relationships among important physical variables. Novel climates are expected to become more common globally under anthropogenic climate change (Wolkovich, Cook, McLauchlan, & Davies, 2014) and have the potential to produce surprising ecological responses, poorly constrained by existing ecological understanding (Williams & Jackson, 2007). We envision that joint analysis of physical and biological data (such as demonstrated using EOF regression) provides a basis for comparing fit and forecasting skill for a range of simple (stationary) to complex (time-varying) linkages between these two model components.

the ACLIM project, which contributed to developing the Bering10K ROMS projection used here. We also thank the EcoFOCI program for prior support of the eastern Bering Sea ROMS models, and K. Kristensen, H. Skaug, and the developers of Template Model Builder without whom this model would not be practical to implement. The authors have no conflicts of interest to declare.

## ORCID

*James T. Thorson* 🆔 https://orcid.org/0000-0001-7415-1010
*Wei Cheng* 🆔 https://orcid.org/0000-0002-5892-2177
*Albert J. Hermann* 🆔 https://orcid.org/0000-0002-0253-7464
*James N. Ianelli* 🆔 https://orcid.org/0000-0002-7170-8677
*Michael A. Litzow* 🆔 https://orcid.org/0000-0003-1611-4881
*Cecilia A. O'Leary* 🆔 https://orcid.org/0000-0002-1737-9294
*Grant G. Thompson* 🆔 https://orcid.org/0000-0002-3240-6705

## REFERENCES

Banerjee, S, Gelfand, A. E, Finley, A. O, & Sang, H (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4), 825–848. https://doi.org/10.1111/j.1467-9868.2008.00663.x

Bindoff, N. L., Cheung, W. W. L., Kairo, J. G., Arstegui, J., Guinder, V. A., Hallberg, R., ... Williamson, P. (2019). Changing ocean, marine ecosystems, and dependent communities. In H. O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, ... N. M. Weyer (Eds.), *IPCC special report on the ocean and cryosphere in a changing climate*. In press.

Brander, K (2010). Impacts of climate change on fisheries. *Journal of Marine Systems*, 79(3), 389–402. https://doi.org/10.1016/j.jmarsys.2008.12.015

Brooks, E. N, & Deroba, J. J (2015). When "data" are not data: The pitfalls of post hoc analyses that use stock assessment model output. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(4), 634–641. https://doi.org/10.1139/cjfas-2014-0231

Conn, P. B, Thorson, J. T, & Johnson, D. S (2017). Confronting preferential sampling when analysing population distributions: Diagnosis and model-based triage. *Methods in Ecology and Evolution*, 8(11), 1535–1546. https://doi.org/10.1111/2041-210X.12803

Cushing, D. H (1990). Plankton production and year-class strength in fish populations: An update of the match/mismatch hypothesis. *Advances in Marine Biology*, 26, 249–293

Deyle, E. R, May, R. M, Munch, S. B, & Sugihara, G (2016). Tracking and forecasting ecosystem interactions in real time. *Proceedings of the Royal Society B*, 283(1822), 20152258. https://doi.org/10.1098/rspb.2015.2258

Dickey-Collas, M, Payne, M. R, Trenkel, V. M, & Nash, R. D. M (2014). Hazard warning: Model misuse ahead. *ICES Journal of Marine Science*, 71(8), 2300–2306. https://doi.org/10.1093/icesjms/fst215

Dormann, C. F, Elith, J, Bacher, S, Buchmann, C, Carl, G, Carré, G, ... Lautenbach, S (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Edwards, A. M, & Auger-Méthé, M (2019). Some guidance on using mathematical notation in ecology. *Methods in Ecology and Evolution*, 10(1), 92–99. https://doi.org/10.1111/2041-210X.13105

Finney, B. P, Alheit, J, Emeis, K.-C, Field, D. B, Gutiérrez, D, & Struck, U (2010). Paleoecological studies on variability in marine fish populations: A long-term perspective on the impacts of climatic change on marine ecosystems. *Journal of Marine Systems*, 79(3), 316–326 https://doi.org/10.1016/j.jmarsys.2008.12.010

Foss-Grant, A. P, Zipkin, E. F, Thorson, J. T, Jensen, O. P, & Fagan, W. F (2016). Hierarchical analysis of taxonomic variation in intraspecific competition across fish species. *Ecology*, 97(7), 1724–1734. https://doi.org/10.1890/15-0733.1

Gleick, P. H (1986). Methods for evaluating the regional hydrologic impacts of global climatic changes. *Journal of Hydrology*, 88(1), 97–116. https://doi.org/10.1016/0022-1694(86)90199-X

Grimmer, M (1963). The space-filtering of monthly surface temperature anomaly data in terms of pattern, using empirical orthogonal functions. *Quarterly Journal of the Royal Meteorological Society*, 89(381), 395–408. https://doi.org/10.1002/qj.49708938111

Hermann, A. J, Gibson, G. A, Cheng, W, Ortiz, I, Aydin, K, Wang, M, ... Holsman, K. K (2019). Projected biophysical conditions of the Bering Sea to 2100 under multiple emission scenarios. *ICES Journal of Marine Science*, 76(5), 1280–1304. https://doi.org/10.1093/icesjms/fsz043

Hjort, J (1926). Fluctuations in the year classes of important food fishes. *ICES Journal of Marine Science*, 1(1), 5–38 https://doi.org/10.1093/icesjms/1.1.5

Holsman, K. K, Essington, T, Miller, T. J, Koen-Alonso, M, & Stockhausen, W. J (2012). Comparative analysis of cod and herring production dynamics across 13 northern hemisphere marine ecosystems. *Marine Ecology Progress Series*, 459, 231–246. https://doi.org/10.3354/meps09765

Huang, B, Thorne, P. W, Banzon, V. F, Boyer, T, Chepurin, G, Lawrimore, J. H, & Zhang, H. M (2017). NOAA extended reconstructed sea surface temperature (ERSST), version 5. *NOAA National Centers for Environmental Information*

Karp, M. A, Peterson, J. O, Lynch, P. D, Griffis, R. B, Adams, C. F, Arnold, W. S, ... Link, J. S (2019). Accounting for shifting distributions and changing productivity in the development of scientific advice for fishery management. *ICES Journal of Marine Science*, 76(5), 1305–1315. https://doi.org/10.1093/icesjms/fsz048

Kearney, K., Hermann, A., Cheng, W., Ortiz, I., & Aydin, K. (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. *Geoscientific Model Development*, 13(2), 597–650. https://doi.org/10.5194/gmd-13-597-2020

Kidson, J. W (1975). Tropical eigenvector analysis and the southern oscillation. *Monthly Weather Review*, 103(3), 187–196. https://doi.org/10.1175/1520-0493(1975)103<0187:TEAATS>2.0.CO;2

Latimer, A. M, Banerjee, S, Sang Jr, H, Mosher, E. S, & Silander Jr, J. A (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(2), 144–154 https://doi.org/10.1111/j.1461-0248.2008.01270.x

Laurel, B. J, Spencer, M, Iseri, P, & Copeman, L. A (2016). Temperature-dependent growth and behavior of juvenile Arctic cod (*Boreogadus saida*) and co-occurring North Pacific gadids. *Polar Biology*, 39(6), 1127–1135. https://doi.org/10.1007/s00300-015-1761-5

Lauth, R. R., & Conner, J. (2014). *Results of the 2011 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate fauna* (NOAA Technical Memorandum NMFS-AFSC-266; p. 185). Alaska Fisheries Science Center. Retrieved from http://www.afsc.noaa.gov/Publications/AFSC-TM/NOAA-TM-AFSC-266.pdf

Lehodey, P, Alheit, J, Barange, M, Baumgartner, T, Beaugrand, G, Drinkwater, K, … Werner, F (2006). Climate variability, fish, and fisheries. *Journal of Climate*, 19(20), 5009–5030. https://doi.org/10.1175/JCLI3898.1

Lehodey, P, Bertignac, M, Hampton, J, Lewis, A, & Picaut, J (1997). El Niño Southern Oscillation and tuna in the western Pacific. *Nature*, 389(6652), 715–718

Ludwig, D, & Walters, C. J (1981). Measurement errors and uncertainty in parameter estimates for stock and recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 38(6), 711–720 https://doi.org/10.1139/f81-094

Mantua, N. J, Hare, S. R, Zhang, Y, Wallace, J. M, & Francis, R. C (1997). A pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, 78(6), 1069–1080. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2

Maunder, M. N, & Thorson, J. T (2019). Modeling temporal variation in recruitment in fisheries stock assessment: A review of theory and practice. *Fisheries Research*, 217, 71–86. https://doi.org/10.1016/j.fishres.2018.12.014

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(12), 1167–1183. https://doi.org/10.1111/ecog.00839

Myers, R. A (1998). When do environment–recruitment correlations work? *Reviews in Fish Biology and Fisheries*, 8(3), 285–305

Myers, R. A (2001). Stock and recruitment: Generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. *ICES Journal of Marine Science: Journal Du Conseil*, 58(5), 937–951. https://doi.org/10.1006/jmsc.2001.1109

Niu, S, Luo, Y, Dietze, M. C, Keenan, T. F, Shi, Z, Li, J, & Iii, F. S. C (2014). The role of data assimilation in predictive ecology. *Ecosphere*, 5(5), art65. https://doi.org/10.1890/ES13-00273.1

O'Leary, C. A, Miller, T. J, Thorson, J. T, & Nye, J. A (2018). Understanding historical summer flounder (*Paralichthys dentatus*) abundance patterns through the incorporation of oceanography-dependent vital rates in Bayesian hierarchical models. *Canadian Journal of Fisheries and Aquatic Sciences*, 76(8), 1275–1294. https://doi.org/10.1139/cjfas-2018-0092

Payne, M. R, Hobday, A. J, MacKenzie, B. R, Tommasi, D, Dempsey, D. P, Fässler, S. M. M, … Villarino, E (2017). Lessons from the first generation of marine ecological forecast products. *Frontiers in Marine Science*, 4, 1–15. https://doi.org/10.3389/fmars.2017.00289

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(11), 1655–1666. https://doi.org/10.1139/cjfas-2013-0508

Sigler, M. F, Stabeno, P. J, Eisner, L. B, Napp, J. M, & Mueter, F. J (2014). Spring and fall phytoplankton blooms in a productive subarctic ecosystem, the eastern Bering Sea, during 1995–2011. *Deep Sea Research Part II: Topical Studies in Oceanography*, 109, 71–83. https://doi.org/10.1016/j.dsr2.2013.12.007

Smith, T. D (2007). *Scaling fisheries: The science of measuring the effects of fishing, 1855–1955* (1st ed.). Cambridge, UK: Cambridge University Press.

Stabeno, P. J, Farley Jr, E. V, Kachel, N. B, Moore, S, Mordy, C. W, Napp, J. M, … Sigler, M. F (2012). A comparison of the physics of the northern and southern shelves of the eastern Bering Sea and some implications for the ecosystem. *Deep Sea Research Part II: Topical Studies in Oceanography*, 65–70, 14–30. https://doi.org/10.1016/j.dsr2.2012.02.019

Szuwalski, C. S, Britten, G. L, Licandeo, R, Amoroso, R. O, Hilborn, R, & Walters, C (2019). Global forage fish recruitment dynamics: A comparison of methods, time-variation, and reverse causality. *Fisheries Research*, 214, 56–64. https://doi.org/10.1016/j.fishres.2019.01.007

Thorson, J. T (2019). Forecast skill for predicting distribution shifts: A retrospective experiment for marine fishes in the Eastern Bering Sea. *Fish and Fisheries*, 20(1), 159–173. https://doi.org/10.1111/faf.12330

Thorson, J. T, Ciannelli, L, & Litzow, M. A (2020). Defining indices of ecosystem variability using biological samples of fish communities: A generalization of empirical orthogonal functions. *Progress in Oceanography*, 181, 102244. https://doi.org/10.1016/j.pocean.2019.102244

Thorson, J. T, Jensen, O. P, & Zipkin, E. F (2014). How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(7), 973–983. https://doi.org/10.1139/cjfas-2013-0645

Thorson, J. T, Scheuerell, M. D, Shelton, A. O, See, K. E, Skaug, H. J, & Kristensen, K (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, 6(6), 627–637. https://doi.org/10.1111/2041-210X.12359

Tobler, W. R (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234–240 https://doi.org/10.2307/143141

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

Williams, J. W, & Jackson, S. T (2007). Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, 5(9), 475–482. https://doi.org/10.1890/070037

Wolkovich, E. M, Cook, B. I, McLauchlan, K. K, & Davies, T. J (2014). Temporal ecology in the Anthropocene. *Ecology Letters*, 17(11), 1365–1379. https://doi.org/10.1111/ele.12353

Zuur, A. F, Fryer, R. J, Jolliffe, I. T, Dekker, R, & Beukema, J. J (2003). Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics*, 14(7), 665–685 https://doi.org/10.1002/env.611

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Thorson JT, Cheng W, Hermann A, et al. Empirical orthogonal function regression: Linking population biology to spatial varying environmental conditions using climate projections. *Glob Change Biol*. 2020;26:4638–4649. https://doi.org/10.1111/gcb.15149