Defining indices of ecosystem variability using biological samples of fish communities: a generalization of empirical orthogonal functions

James T. Thorson[1*], Lorenzo Ciannelli[2], Michael A. Litzow[3]

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

[2] College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Kodiak, AK 99615, USA

[3] College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, 97331

* Corresponding author

James.Thorson@noaa.gov

Keywords: Empirical orthogonal function; sea-ice extent; eastern Bering Sea; bottom trawl; vector autoregressive spatio-temporal model

Abstract:

Multivariate data reduction techniques are widely used to describe modes of variability in atmospheric and oceanographic conditions for the world's oceans. Dominant modes of variability such as the Pacific Decadal Oscillation (PDO) are typically defined as a statistical summary of physical measurements, and include both principle components representing modes of variability over time, and an empirical orthogonal function (EOF) giving the spatial pattern associated with a positive or negative phase for each mode. Typically, these indices are compared with biological conditions to describe or predict physical drivers of ecological dynamics. In some circumstances, however, it may instead be useful to apply EOF analysis directly to biological measurements, estimating indices of biological variability as well as maps of biological response associated with each index. We therefore develop a generalization of EOF analysis that can be applied directly to multispecies biological samples using a multivariate spatio-temporal model. These biologically derived indices can then be compared with relevant indices derived from physical data, or used as covariates in spatially-varying coefficient models. We first show that a spatio-temporal model can replicate previous EOF estimates of the PDO and North Pacific Gyre Oscillation. We then identify three axes of variability in the eastern Bering Sea using biomass-sampling data for fourteen bottom-associated fishes and decapod crustaceans from 1982-2017. The first axis represents habitat preferences that are stable over time, and the second represents a multi-decadal trend in distribution for most species; for example, showing an increasing density for Alaska skate and arrowtooth flounder in the middle and inner domain. Finally, the third axis shows high interannual variability from 1982-1998 switching to multiyear stanzas from 1999-2017 and is highly correlated (0.87) with the extent of the cold bottom temperatures in this region and associated impacts on Alaska pollock and Pacific cod. These

41 axes represent ecological dynamics for adult fishes and therefore integrate the impact of bottom-
42 up and top-down processes, and they also confirm the importance of cold-pool extent for fish
43 distribution in the Bering Sea while visualizing its varied impact on individual species.
44 Moreover, this spatio-temporal approach allows oceanographers to define annual indices
45 representing modes of variability in diverse biological communities from widely available field-
46 sampling data.
47

Introduction

Oceanographers are confronted with a tremendous challenge in summarizing multivariate physical and biological processes into a tractable number of dominant patterns, which can then be readily communicated to other fields (e.g., fisheries scientists), used as covariates in climate models, or used in many other ways. One approach to this challenge is to identify dominant axes of variability for a given process, and then to represent the process as an index that varies over time, combined with a loadings map expressing the spatial pattern associated with the index. Familiar examples include the Pacific Decadal Oscillation (PDO; Mantua et al., 1997) and North Pacific Gyre Oscillation (NPGO; Di Lorenzo et al., 2008). These indices capture patterns of basin-scale variability that play leading roles in the dynamics of regional ecosystems (Schwing et al., 2010), and also tend to capture variability in a suite of cross-correlated, ecologically important processes (Stenseth et al., 2003), making them valuable tools for summarizing climate effects on marine biota. For example, the importance of the PDO was established via comparison with records of salmon returns in the Pacific Ocean (Mantua et al., 1997), while the importance of the NPGO was demonstrated via its correlation with nutrients and chlorophyll concentrations in southern California (NPGO; Di Lorenzo et al., 2008). These ocean climate indices are typically referred to as principal components (PCs) for the temporal index and empirical orthogonal functions (EOFs) for the loadings (e.g., Trenberth et al., 2014). Here we broadly use the term "EOF analysis" to refer to the algorithm generating both the temporal index and the spatial map associated with it, while recognizing that different authors use different algorithms for individual steps in the EOF analysis (e.g., PCA vs. factor analysis for generating the temporal index).

There have been several applications of EOF analysis to biological spatio-temporal data in the last decade (e.g., Morfin et al., 2012; Marshall et al., 2016). However, previous applications of data reduction techniques applied to biological data involve either the aggregation across space and time so to obtain regional averages of multi-species biomass (e.g., PCA applied to annual indices of species biomass, Planque and Arneberg (2018)), or aggregation across species so to obtain single-species measurement of biomass across multiple sampling sites (e.g., EOF analysis of spatio-temporal data from single species). These data manipulations inevitably reduce the scope of ecological inference, sacrificing either small-scale spatiotemporal dynamics or the ability to generalize the effect of climate across multiple taxa (Puerta et al., 2019).

In particular, we identify three technical hurdles when applying EOF analysis directly to biological sampling data: (1) biomass-sampling data has many zeros with few extreme values (a highly skewed "dust bunny distribution" *sensu* McCune and Root (2015)) which is poorly represented using a normal distribution; (2) biological samples often arise from spatially unbalanced sampling, which complicates the creation of spatial maps in unsampled areas or when fitting to spatially unbalanced sampling programs; and (3) biological sampling is often available for multiple species, and therefore EOF analysis must be generalized for multivariate data sets, where each variable has both temporal and spatial expression.

Despite the profusion of research using EOFs to characterize physical oceanographic conditions, it is not always clear how to relate these indices to biological processes. This commonly involves a multi-step workflow, where physical variability is first summarized with EOF analysis, and the resulting index is then used as a covariate in a separate statistical model linking physical oceanography to biological or social outcomes. Unfortunately, many published relationships between physical conditions and biological responses in ocean environments either

93 break down over time or have poor skill when forecasting (Myers, 1998; Thorson, 2019a). We
94 note two ways that this multi-step workflow may contribute to poor out-of-sample predictive
95 skill for statistical physics-biology relationships. First, climate indices derived from EOF
96 analysis summarize physical dynamics across a large spatial domain. These regional indices are
97 then correlated with local conditions, and biological responses are then correlated with local
98 conditions for each individual population (Newman et al., 2016; Stenseth et al., 2003; Wills et
99 al., 2018). However, the correlation between regional indices and local physical conditions can
100 change over time, thus changing their correlation with local biological responses (Litzow et al.,
101 2018; Newman et al., 2016). Second, the solution of the EOF analysis is defined such that the
102 first axis explains the most variance in the underlying data, the second axis explains the second-
103 most, etc. However, these definitions are fixed in time while different axes of physical variation
104 may have a larger impact on biological during some conditions than others, e.g., physical
105 variation associated with juvenile production may be more important after fishing has reduced
106 the age-structure of a population (Hsieh et al., 2006).
107 Given these difficulties when using oceanographic indices of physical habitat to describe
108 biological responses, we see a useful role for indices estimated directly from multivariate
109 biological variables. In particular, recent improvements in statistical computation and spatial
110 statistics have allowed the growth of multivariate spatio-temporal models (Clark et al., 2014;
111 Latimer et al., 2009; Ovaskainen et al., 2017). For example, spatial factor analysis (Thorson et
112 al., 2015b) could be used to generalize EOF analysis using biomass samples for multiple species
113 simultaneously, and therefore could represent indices of variation in the ecological dynamics that
114 arise from both bottom-up and top-down processes. We propose that these approaches would be
115 useful for compressing one or more variables to a single (or few) easily visualized time series.

Indices derived from biological sampling data could then be correlated with similar indices estimated from physical data to explore physics-biology linkages, as a descriptive summary of ecological dynamics (e.g., McClatchie et al., 2018), or used directly as covariates in other biological models (e.g., O'Leary et al., 2018).

In this study, we develop a method to estimate dominant modes of variability (i.e., one or more indices, each associated with a map showing spatial patterns in the positive phase of a given index) that can be applied to noisy, multispecies field samples of biological variables. We then demonstrate that this approach generalizes conventional EOF analysis by replicating estimates of PDO and the NPGO using North Pacific sea surface temperature fields. Finally, we demonstrate the approach using data for fourteen bottom-associated fish and decapod species in the eastern Bering Sea. Many studies in this system have linked the spatial distribution of fishes to the location and spatial extent of cold near-bottom temperatures (e.g., Wyllie-Echeverria and Wooster, 1998; Baker and Hollowed, 2014), although long-term trends in distribution appear to be independent of cold-pool extent for some species (Mueter and Litzow, 2008). Similarly, our method estimates an index of ecosystem variability that is highly correlated with the spatial extent of cold near-bottom waters (termed the "cold pool"), and therefore corroborates the important role and estimates the spatially varying effect of coupled winter ice cover and summer bottom temperature in that region over the past 36 years.

Methods

A brief history of Empirical Orthogonal Function (EOF) analysis in physical oceanography

Climatologists, meteorologists, and oceanographers have used empirical orthogonal functions to characterize dominant modes of variability in physical ocean conditions for over fifty years. Grimmer (1963) applied factor analysis to sea surface temperature anomalies in the North

Atlantic to show that 80% of variance can be explained by a small number of latent variables. Kidson (1975a) introduced a similar method using principal components analysis (PCA) applied to monthly average precipitation and sea surface pressure to similarly show that a reduced set of variables could explain >75% of the original variance. These studies established the common practice wherein a multivariate statistical techniques (PCA or factor analysis) is applied to spatially replicated measurements of a physical variable at multiple times (e.g., years or months), where a small number of dominant axes can be used to explain the large portion of variance in the original process.

Empirical orthogonal functions have subsequently been used to define or analyze several of the most widely-known global and regional processes in oceanography. Kidson (1975b) used EOF to measure the Southern Oscillation, a pattern in equatorial surface pressure and precipitation that was originally described by Walker (1924). More recently, Mantua et al. (1997) used EOF to define the PDO as the dominant mode of variability in sea surface temperature (SST) in the North Pacific, and the second mode was further explored and interpreted by Bond et al. (2003) and subsequently called the "Victoria mode" (Ding et al., 2015). Similarly, Di Lorenzo et al. (2008) defined the NPGO as the second mode of variability for sea surface height (SSH) anomalies, and the NPGO is correlated with the Victoria mode via the high correlation between SST and SSH.

The PDO and NPGO have seen broad use in biological oceanography due to the correlation between these physical indices and biological production that underlies changes in fishery productivity (Di Lorenzo et al., 2008; Mantua et al., 1997). For example, the importance of the PDO was originally demonstrated based on cycles in salmon productivity between Alaska and Oregon/Washington stocks (Mantua et al., 1997). However, the correlation between the PDO

and salmon productivity has declined since the description of the PDO, apparently due to the changing associations between the PDO and localized physical processes that underlie salmon productivity for individual salmon stocks (Litzow et al., 2018).

Finally, we note ongoing research that defines EOFs in the context of a statistical model that separately estimates measurement errors from physical variation. For example, Grimmer (1963) used factor analysis to separate measurement and process errors, and Ghil et al. (1981) subsequently defined a multivariate Kalman filter that explicitly models covariation in physical dynamics. This Kalman-filter interpretation of EOF was subsequently extended by Wikle and Cressie (1999), who introduced a descriptive spatial process wherein the physical process is more similar at nearby than at distant locations. Subsequently, these "spatio-temporal" models have rapidly developed due to improvements in statistical and computational techniques (e.g., Lindgren et al., 2011), and now are also widely used in ecology and fisheries science (Ovaskainen et al., 2017; Thorson, 2019b). To our knowledge, however, this statistical generalization of EOF analysis to generate the index and the spatial map simultaneously has not been applied to multi-guild biological data (e.g., abundance or biomass for species in a community) while accounting for variable dynamics across both space and time.

General approach

We develop a method to estimate one or more ecosystem indices based on biomass-sampling data for multiple species. Each index f includes two components:

1. a spatial map $\mathcal{X}_{f,c}(s)$ representing whether a given location s has a positive or negative value (as well as the magnitude of that value) during a "positive phase" of the index; and
2. a time-series $\lambda_f(t)$ indicating whether a given time t has a positive phase ($\lambda_f(t) > 0$) or negative phase ($\lambda_f(t) < 0$), as well as whether a given year has a weak magnitude ($\lambda_f(t)$

within the 25% and 75% quantiles for λ_f) or a strong magnitude ($\lambda_f(t)$ outside the 25% and 75% quantiles).

Importantly, the map $\mathcal{X}_{f,c}(s)$ associated with each time series $\lambda_f(t)$ differs for each category c of n_c modeled categories, and this allows our time series to represent multivariate data, for example, multispecies biomass samples from bottom trawl surveys. In the following, we focus on annual variation (i.e., t indexes different years), but the process could instead represent other time intervals (weekly, monthly, etc.).

To estimate these ecosystem indices, we begin by defining a predictor variable $\tilde{\mathcal{Y}}_{c,t}(s)$ that includes the net effect of all estimated oceanographic indices. Each predictor variable is a linear combination of n_f oceanographic indices:

$$g\left(\underbrace{\tilde{\mathcal{Y}}_{c,t}(s)}_{predictor\ variable}\right) = \underbrace{\beta_{c,t}}_{intercepts} + \underbrace{\sum_{f=1}^{n_f} \overbrace{\lambda_f(t)}^{Index\ phase}\ \overbrace{\mathcal{X}_{f,c}(s)}^{Index\ map}}_{Net\ effect\ of\ indices} \quad (1)$$

where $g(\tilde{y})$ is a link function transforming the linear predictor to response $\tilde{y}$ and $\beta_{c,t}$ are intercepts that vary among categories and times. This predictor variable is then estimated by minimizing the difference between it and measurements $\mathcal{Y}_{c,t}(s)$ of each response variable:

$$\mathcal{Y}_{c,t}(s) \sim h\left(\widetilde{\mathcal{Y}_{c,t}}(s), \dots\right) \quad (2)$$

where h is a probability distribution function for measurements $\mathcal{Y}_{c,t}(s)$ given their predicted values $\widetilde{\mathcal{Y}_{c,t}}(s)$. For a discussion of how this general model relates to previously developed spatio-temporal models, please see Appendix 1.

Demonstrating the similarity to Empirical Orthogonal Function (EOF) analysis

We first seek to show that this approach can generalize the EOF analysis that is widely used in physical oceanography to generate indices. Published studies have implemented EOF using

either principle components analysis (PCA) or factor analysis (FA), and both involve a matrix $Y_t(s)$ of measurements of a physical variable at each location and time. Anomalies from long-term climate are then typically calculated, $Y_t^*(s) = Y_t(s) - \frac{1}{n_t}\sum_{t=1}^{n_t} Y_t(s)$, and these anomalies are used to calculate the sample covariance $V_{t,t}$ for anomalies between any pair of times. The analyst then applies an eigendecomposition (for PCA) or minimization algorithm (FA) to identify a set of n_f orthogonal axes of covariation, ranked from most important to least important, such that the first few axes explain the majority of covariation. These axes are then treated as the index $\lambda_f(t)$, and the spatial map $x_f(s)$ associated with each index is calculated as either the correlation or regression of $\lambda_f(t)$ and $y_f(s,t)$.

We replicate a univariate EOF analysis by simplifying the general model in four ways:

1. Eliminate notation for multiple categories such that it is applied to a single response;
2. Use an identity link function;
3. Use a normal distribution for observations; and
4. Replace functions with matrices, i.e., $\mathcal{X}_{f,c}(s)$ with a matrix **X** representing the predicted value of each oceanographic variable $x^*(s_i, f)$ at the location for each sample i for each index, and $\tilde{\mathcal{Y}}_{c,t}(s)$ with $\tilde{y}_i$ representing the prediction for that sample i.

These three changes result in the following model:

$$\tilde{y}_i = \beta(t_i) + \sum_{f=1}^{n_f} \lambda_f(t_i) x^*(s_i, f) \quad (3a)$$

$$y_i \sim Normal(\tilde{y}_i, \sigma^2) \quad (3b)$$

where σ^2 is the variance of measurement errors, which is minimized by explaining variation in the response $y(s,t)$ to estimated indices, and $x^*(s_i, f)$ is calculated from a predictive-process model of spatial variation. We note that we do not center the data $y(s,t)$ prior to analysis using

224 this generalized model, so the first axis in this generalized model corresponds to persistent
225 spatial differences that would otherwise be eliminated by centering the data. This is different
226 from conventional EOF analysis (which does center the data to calculate anomalies prior to
227 analysis), and this difference means that the 2nd axis from the generalized model is similar to the
228 1st axis from conventional EOF, the 3rd axis is similar to the 2nd from conventional EOF, etc.
229 However, centering the data does not extend to other common forms of data analysis (e.g.,
230 generalized linear models) so not centering the data is important for subsequent generalizations.
231 Also differing from conventional EOF, we estimate a separate intercept β_t for every year, so that
232 this term captures interannual variability in $y(s,t)$ among years (i.e., the increasing trend in
233 surface temperatures due to climate forcing). We estimate a separate intercept for each year to
234 match model specification for the multi-species extension to biological sampling data, as
235 justified below, and future applications could easily specify a model where this intercept is
236 constant across years (to be more similar with conventional EOF analysis). We then compare
237 results with public values for the PDO (obtained Nov. 6, 2018 from
238 http://research.jisao.washington.edu/pdo/PDO.latest.txt) and NPGO (obtained Nov. 6, 2018 from
239 http://www.o3d.org/npgo/).
240 *Extension to multi-species biomass-sampling data*
241 We also seek to show how this approach can include biomass-sampling data for multiple species,
242 such as are widely available worldwide from resource surveys of fish stocks. For multivariate
243 biomass-sampling data we define a more complicated sampling process, which involves several
244 modifications to the general model:
245 1. Use multiple categories, each with an independent intercept for each species and year;

2. Use a Poisson-link delta-model (Thorson 2017) where the linear predictor is associated with numbers density n, where encounter probability p_i for sample i is derived as a complementary log-log link from numbers density $p_i = 1 - \exp(-n_i)$ and expected biomass when encountered r_i is defined such that it is proportional to numbers density, $r_i = \frac{n_i}{p_i} w(c_i, t_i)$, where $w(c_i, t_i)$ is the biomass per individual, which is estimated separately for each category and year; and
3. Minimize the negative log-likelihood for encounters and sampled biomass for each species.

These changes result in the following model:

$$\log(n_i) = \beta(c_i, t_i) + \sum_{f=1}^{n_f} \lambda_f(t_i) x^*(s_i, c_i, f) \quad \text{4a}$$

$$\Pr\big(B = b(s_i, c_i, t_i)\big) = \begin{cases} 1 - p_i & \text{if } B = 0 \\ Lognormal(\log(r_i), \sigma_t^2) & \text{if } B > 0 \end{cases} \quad \text{4b}$$

Where $\beta(c, t)$ is again an intercept for every category and time, Eq. 4b represents a delta-model distribution where $1 - p_i$ is the probability mass associated with $B = 0$, $Lognormal(a, b)$ is a lognormal probability density function with logmean a and log-variance b, and σ_t^2 is the residual estimated log-variance in positive catch rates. We specify a separate intercept for every category and time because, in our experience, total abundance for marine fishes often varies substantially among years due endogenous biological processes (e.g., variable production of juveniles; Thorson et al. (2015a)) and the annually varying intercept "controls" for this variation such that remaining model explain shifts in distribution rather than total abundance. This model includes a log-link (Eq. 4a) such that variation a 0.01 increase in $\lambda_f(t_i) x^*(s_i, c_i, f)$ corresponds to an approximately 1% increase in expected biomass. This relative scale implies that all coefficients representing spatial and temporal variation (i.e., the right-hand side of Eq. 4a) are dimensionless.

In the introduction, we noted three difficulties with applying EOF to biological sampling data: (1) many zeros and a skewed distribution; (2) spatially unbalanced sampling; and (3) multiple response variables. These difficulties are addressed by (1) applying a delta-model with a linear predictor that is shared among years and species; (2) projecting the linear predictor from n_s knots to any possible location within the modeled spatial domain, including locations with missing or spatially misaligned data; and (3) associating each mode of variability $\lambda_f(t)$ with a map representing the impact $x^*(s,c,f)$ of a positive phase for each of n_c modeled variables at any location s within a defined spatial domain. In particular, accounting for spatially misaligned data allows EOF to be applied to multiple sampling programs, operating at different locations or over different spatial domains, although we do not explore the idea further here.

Parameter estimation

Parameters for both configurations of the general model can be estimated using a publicly available R package VAST for vector autoregressive spatio-temporal models (VAST; Thorson and Barnett, 2017), using release number 3.1.0 (https://github.com/James-Thorson-NOAA/VAST). This package estimates spatial variables $x(s,c,f)$ for n_s "knots" as random effects following a Gaussian Markov random field, uses a predictive process formulation to interpolate the value of $x^*(s_i,c_i,f)$ for the location s_i and category c_i of sample i given $x(s,c,f)$, and models the correlation in $x(s,c,f)$ between any two knots using a stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011) which approximates a Matérn correlation function (see Appendix 2 for details). Fixed effects are then estimated using maximum likelihood (ML) techniques while approximating the marginal likelihood using the Laplace approximation (Skaug and Fournier, 2006), and efficiently identifying the ML estimates using automatic differentiation (Fournier et al., 2012) as implemented using the TMB package in

R (Kristensen et al., 2016). Further details regarding VAST can be found elsewhere (Thorson 2019).

Factor-analysis models generally require some constraints on loadings matrix $\lambda_f(t)$ to ensure that the model is identifiable. We follow previous practice in fixing $\lambda_f(t) = 0$ for all $f > t$ (Thorson et al., 2015b; Zuur et al., 2003), but then rotate results to ensure that they are interpretable similarly to principle components analysis (PCA). In particular, we define a rotation matrix $\mathbf{R}$ such that $\mathbf{\Lambda R}$ has columns identical to the eigenvectors of $\mathbf{\Lambda}^t\mathbf{\Lambda}$, and then define $\mathbf{\Lambda R}$ as the climate indices and $\mathbf{RX}(c)$ as the map for each category c. We specifically use a "PCA rotation" (Thorson, 2019b; Thorson et al., 2016a), which maximizes the variance for each axis in sequential order, and this differs from varimax rotation which would instead associate each mode of variability with a minimal subset of species.

Validating EOF application using sea surface temperature

Having defined a statistical model that generalizes a conventional EOF analysis, we seek to validate that it yields estimates that are similar to an EOF analysis. We therefore download monthly average sea surface temperature for every 2° by 2° grid cell within 20°-60°N and 132°-250°W for every month from Jan. 1950 through Jan. 2018 from the NOAA Extended Reconstructed Sea Surface Temperature (v.5) product (Huang et al., 2017). We analyze these data with Eq. 3a-3b using three indices, and hypothesize that the 1st index will represent spatial differences in average temperature, the 2nd will be correlated with the Pacific Decadal Oscillation (PDO), and the 3rd will be correlated with the North Pacific Gyre Oscillation (NPGO). However, we note that small differences with the PDO will likely arise because we are analyzing only January SST, while the conventional PDO index is calculated from monthly SST measurements. We also note that differences are likely to be greater for the NPGO because the conventional

311 NPGO is calculated from sea surface height (SSH), and while SST and SSH are often correlated
312 they are not guaranteed to be perfectly correlated.
313 *Estimating an ecosystem index from bottom-trawl data in the eastern Bering Sea*
314 We also seek to demonstrate the benefits of estimating oceanographic indices directly from
315 multispecies biomass-sampling data. To do so, we download bottom trawl sampling data from
316 the eastern Bering Sea bottom trawl survey (Lauth and Conner, 2016) for fourteen commonly
317 occurring bottom-associated fish and crab species. We then fit Eq. 4a-4b to these data while
318 again estimating three oceanographic indices. We hypothesize that the 1st will represent species-
319 specific habitat preferences that are stable over time, and do not have strong *a priori* hypotheses
320 regarding the remaining two indices.

321 ## Results

322 *Validating EOF application using sea surface temperature*
323 Visualization of the three dominant indices fitted to January average sea surface temperature
324 (Fig. 1) confirms that the spatio-temporal model uses the 1st axis to represent persistent spatial
325 differences in temperature, and the 2nd axis accurately reproduces the annual PDO index with a
326 correlation of 0.92 with the PDO calculated as the EOF of these data. The spatial map associated
327 with the PDO index confirms that years with a positive PDO phase have warmer temperatures
328 inshore off the coast of North America and cooler temperatures in the central North Pacific.
329 Similarly, the 3rd index is correlated with the NPGO, although the correlation (0.52) is
330 substantially weaker than for the PDO. The map associated with this index is orthogonal to the
331 PDO map, and shows that a positive phase for the NPGO is associated with decreased water
332 temperatures in the California Current from California southward, and elevated temperatures in
333 the northern Bering Sea. The maps for PDO and NPGO are correlated with published estimates

and, as expected, the annually varying intercepts ($\beta(t)$ in Eq. 3a) show a gradual increase over time representing climate-forcing of surface temperatures in the Pacific Ocean.

Estimating an ecosystem index from bottom-trawl data in the eastern Bering Sea

Visualizing the three indices estimated from bottom-trawl survey data for fourteen bottom-associated species in the eastern Bering Sea (Fig. 2A) shows that these multispecies indices partition temporal variability into stable (1st index), multi-decadal trend (2nd index), or interannual variability (3rd index). As expected, stable habitat preferences (the 1st axis) explains the majority (87%) of total (spatial and spatio-temporal) covariation, while long-term shifts (2nd axis) explains more (9%) than the axis of interannual variability (3rd axis; 4%). For comparison, we also calculate the proportion of temporal covariation explained by each factor; this confirms that the first axis explains only 10% of temporal while axis 2 and 3 both explain considerably more (63% and 28%, respectively).

Visualizing a "habitat preference" index for each species (Fig. 2B, left column) shows expected relationships where, for example, snow crab has a more northward distribution than Tanner crab, arrowtooth flounder has elevated density in the outer domain in the southwestern edge of the survey, and Pacific cod has relatively weak preferences relative to other species. Variation among years in the magnitude of this index (i.e. a higher value in 1995 than 1998) presumably indicates years where species' distribution is more (1995) or less (1998) in-line with their long-term habitat preferences. The "multi-decadal trends" index shows a decrease over the 37 years analyzed here, and its spatial map shows, for example, that it is capturing the long-term decrease in density for Alaska skate and arrowtooth in the outer domain relative to the middle domain, as well as a long-term decrease in Tanner crab near Bristol Bay (Fig. 2B middle column). Finally, the "interannual variability" index shows 1-3 year periods of positive or

negative stanzas from 1982-1998, then changing to 6-8 year stanzas from 1999-2017. A negative phase for this index (Fig. 2B right column) is associated e.g., with increased density of Alaska pollock and Pacific cod on the northern boundary of the eastern Bering Sea survey and decreased density of Tanner crab in this same area (see Appendix 3, Fig. S1 for patterns for other species). These three indices collectively capture several well-documented trends in the eastern Bering Sea, for example, the long-term shift in distribution for arrowtooth from the outer to middle domain (Fig. 3 bottom row). In particular, these indices capture increased density for Alaska pollock and Tanner crab in the northern portion of the survey in 2017 relative to 1982-1996. Similar patterns are also captured when fitting a model that includes independent spatial and spatio-temporal variation for every species (Appendix 3, Fig. S2), so we conclude that these patterns are not artefacts of the EOF model structure but instead are a low-rank representation of patterns found in the bottom-trawl survey data. Finally, we conclude by noting that that the "interannual variability" index is highly correlated (0.74) with the area of the eastern Bering Sea with summer bottom temperatures less than 2° C (Fig. 4), despite the model not fitting to any physical data, where the small standard errors for this estimated index (black whiskers in Fig. 4) suggest that the pattern is generated by signal within available data rather than statistical error introduced by noisy sampling data.

Discussion

In this paper, we have shown that a multivariate spatio-temporal model can be used to apply empirical orthogonal function analysis to multispecies samples of population densities. EOF analysis is widely used to summarize physical processes (e.g., sea surface temperature) but is less-often used to summarize biological processes due to the noisy, zero-inflated, and multivariate nature of common biological sampling. Multispecies marine resource surveys are

routinely collected in standardized designs in many marine ecosystems worldwide, including Europe, North America, East Asia, Southern Africa, and Oceana, and the dominant axes of variability for ecosystem dynamics in these systems can now be summarized using the publicly available software package, *VAST*.

We have also demonstrated our approach using two case studies: temperature in the North Pacific and biomass-samples for bottom-associated fishes and decapods in the eastern Bering Sea. In both examples, the dominant axis represents persistent spatial variation in a given physical or biological variable, and this axis explains 87% (for biological) and 99% (for physical) variance. This portion of variance is not typically quantified during EOF analysis, which is often applied to anomalies and therefore long-term patterns are eliminated prior to analysis. We retain this portion of variance (rather than centering data prior to analysis) so that the same algorithm can be applied to biological sampling data that are best described using a delta-model or other "dust-bunny" distributions (McCune and Root, 2015; Thorson, 2018). The 2nd axis then represents longer-term spatial trends in the biological case-study, in this case represented in part by the shift in distribution of several flatfishes from offshore to onshore areas. Additional trends may be driven by east-west movements (e.g., great sculpin). The distribution shift for arrowtooth flounder has been previously documented (Spencer et al., 2016; Thorson et al., 2016b), and has been attributed to both the extent of the cold pool (Kotwicki and Lauth, 2013) and an interaction between the bottom temperature and population abundance (Ciannelli et al., 2012). The 3rd axis, by contrast, corresponds to the widely documented role of sea-ice and cold-pool extent in the Bering Sea (Hunt et al., 2011; Stabeno et al., 2012). It is comforting that an EOF analysis of biological variables is able to recover this well-documented, bottom-up driver of ecological dynamics. However, the fact that the "cold pool" axis explains less variance

than the "multi-decadal trends" axis underlines the importance of future studies to identifying bottom-up or top-down drivers of long-term spatial redistribution of flatfishes in this region.

We also note a paradox regarding the 3rd axis in the eastern Bering Sea. This axis has an index that closely corresponds to the extent of the cold pool, and fluctuations in cold-pool extent are primarily driven by changes in bottom temperature in the northern portion of the EBS (Appendix 3, Fig. S3) although some cold years (e.g., 1999, 2006-2010, 2012) have large enough cold pool extent that they then decrease bottom temperatures in the southern EBS. However, the species-specific maps associated with the 3rd axis of variability resemble the map of bottom temperatures (Appendix 3, Fig. S3) for some species but not others. In particular, the spatial areas associated with changing bottom temperatures overlap with the 3rd axis maps for several species (e.g., Pacific cod. walleye pollock, Tanner crab) but not others (e.g., yellowfin sole, Alaska plaice, great sculpin). These latter species appear to show changes in spatial distribution between years with large and small cold-pool extent, but their distribution does not necessarily expand as a function of local bottom temperatures. This supports previous research showing that local bottom temperature in isolation is not sufficient to explain or predict future distribution for many EBS species (Litzow, 2017; Thorson et al., 2017b; Thorson, 2019a). In fact, it suggests that distribution for these species may be driven by a mechanism that is correlated with fluctuations in cold-pool extent, but not necessarily bottom temperature itself. This could include changes in the timing of the plankton bloom as driven by timing of ice melt and indirectly related to bottom temperatures (Hunt et al., 2011; Sigler et al., 2014), the timing of ontogenic movement from nearshore to offshore habitats (Nichol et al., 2019) or the opening of thermal gateways across the middle shelf which allow species located in the outer shelf to move toward the inner shelf (Ciannelli and Bailey, 2005). These and other potential mechanisms could

result in a spatial distribution that differs between warm and cold years (and hence is correlated with the cold-pool extent), but where distribution in those years is not necessarily correlated with local bottom temperatures. Alternatively, spatio-temporal dynamics that are in phase with the extent of the cold pool index may also be driven by a change in catchability of the bottom trawl sampling gear. The bottom trawl only samples the lower few meters of the water column during a fixed summer season, and catchability may vary among years either due to variation in gear performance or spatial availability (von Szalay and Somerton, 2005; Nichol et al., 2019). These many processes could result in species catch rates that covary with sea-ice extent but not with localized temperatures, necessitating a flexible approach to modelling region-scale drivers of distribution for these species. Relatedly, these species vary widely in generation time and life-history strategy, such that it is surprising that the majority show a strong response to cold-pool extent (i.e., the 3rd factor). Despite these differences in life-history, our results suggest that these various bottom-associated species all shift their spatial distribution in response to behavioral cues, prey availability, or other mechanism that are strongly correlated with cold-pool extent.

A common approach for examining the effect of large-scale climate fluctuations on the biological productivity of marine organisms is to reduce complex physical dynamics to one or more indices and then correlate these indices with measures of biological production (Stenseth et al., 2003). This approach has contributed to characterizing patterns of climate-productivity covariation in the ocean, but it is challenged by the evolving nature of these relationships over time and space (Litzow et al., 2018). Analysts could also use the statistical generalization of EOF analysis to analyze the link between biology and physics in two new ways. First, as shown here, estimated modes of biological variation can be compared with relevant measurements of physical variables across scales, to provide some interpretation of what physical processes are

449 correlated with leading biological variation. Linking physics and biology in this way has some
450 benefits for interpreting biological modes of variation, while still visualizing the residual
451 biological variation that is not otherwise explained. Importantly, an estimated mode of
452 biological variation may be correlated with a known physical process during one portion of the
453 time-series and not another, and this can be used to indicate when the linkage between these
454 physical and biological processes is nonstationary. Second, biological indices could be extracted
455 and used directly in a secondary modelling framework, e.g., as an annual index in a population-
456 dynamics model (e.g., Schirripa et al., 2009). This second approach allows modes of biological
457 variation to be propagated even when the physical mechanisms driving variation are poorly
458 understood, and therefore represents ecosystem drivers without directly modelling them.
459 In the application to groundfish data, EOF analysis decomposes biological variation into
460 stable patterns (1st mode), multi-decadal trends (2nd mode) and interannual variability (3rd mode).
461 The scale order of these indices may not always be the same, depending on how variability at
462 each scale contributes to the total variability of the biological data set. Similarly, other data sets
463 may lack either longer-term trends or strong internanual variability. Nevertheless, decomposing
464 biological variability across temporal scales may enable a more mechanistic understanding of the
465 effect of climate on biological productivity (Drinkwater et al., 2010; Ottersen et al., 2010).
466 The generalization that we propose here allows applications to data that are not necessarily
467 normally distributed and can handle the zero inflation via application of a delta model. The error
468 structure is flexible to allow temporally correlated measurements and non-stationarity (see
469 Thorson (2019b) for a full description of features available in package *VAST*). The generalization
470 presented here can be used to simultaneously detect patterns of covariation across space and time
471 in multispecies assemblages, and therefore overcome these problems. This ability to provide

simultaneous inference across taxa can provide a mechanistic understanding of bottom up and top down dynamics affecting species abundance/productivity (Cury et al., 2008). We envision future biological oceanography applications for the analyses of meroplankton (McClatchie et al., 2018), remote sensing products for characterization of biogeochemical and ecological oceanic boundaries (Kavanaugh et al., 2014), and zooplankton collections (Colebrook, 1978), in addition to the application demonstrated using survey data for groundfishes.

We see three primary avenues to extend this work, which we recommend for future research:

1. *Spatially varying coefficient models to forecast distribution shifts*: Our analysis for the eastern Bering Sea suggests that many species have a strong but nonlocal response to regional temperatures, as shown by the similarity between the 3rd axis and cold-pool extent. These indices can be included in species distribution modes using a apatially-varying coefficient (Bacheler et al., 2009; Bartolino et al., 2011), and recent research corroborates that a spatially varying response to cold-pool extent improves forecast skill relative to local temperature in isolation (Thorson, In press). We hypothesize that annual indices such as cold-pool extent will be informative about low-frequency variability in species distribution that is currently lacking in long-term forecasts of distribution shift in response to climate.
2. *Joint rank-reduction models of physical drivers and biological responses*: An interesting extension of the generalized EOF analysis used here is to include both physical and biological variables, to make inference on the shared covariance among them and the spatial patterns in which this covariance is dominant (Brown and Fiechter, 2012). The signal of anthropogenic climate change is emerging from the envelope of natural variability in many ecosystems globally (Henson et al., 2017), including the Bering Sea (Walsh et al., 2018). This anthropogenic change makes nonstationary behavior in environmental variables

(changing mean or variance) an increasingly important factor in ecosystem dynamics (Burrows et al., 2011). The spatio-temporal generalization of EOF analysis presented here may be expanded to include both physical and biological variables, to make inferences about shared covariance and to directly infer the role of nonstationary physical dynamics in leading ecosystem patterns (Brown and Fiechter, 2012). In particular, the estimated linkage between physical and biological variables could be useful to estimate the spatial map of physical conditions that is most strongly associated with variation in a given biological variable. In essence, this could be used to identify the physical index that maximizes predictive power for a given species.

3. *Model evaluation via null models and simulation testing*: Finally, we recommend future research regarding the expected performance of applying EOF to multivariate biological samples. Likely model performance could be explored via simulation-testing, where fits to data from multiple ecosystems could be used to simulation new data (potentially with different sampling designs or sample sizes) and the model could then be re-fitted to explore measure model performance. The performance of multivariate spatio-temporal models when estimating "factor-loadings" (analogous to EOF indices in this analysis) and associated spatial maps has been simulation-tested previously (e.g., Thorson et al., 2016a, 2017a, In press), but performance specifically when generalizing EOF analysis remains a useful topic for further testing. Alternatively, researchers could adapt previously developed "null models" (e.g., Planque and Arneberg, 2018) to determine whether the statistical approach to EOF is likely to identify erroneous patterns from autocorrelated noise.

Another issue raised by anthropogenic climate change is the creation of novel patterns of covariance among different ecologically important physical variables, which may produce

nonstationary statistical relationships between environmental drivers and ecological responses, or nonstationary relationships among species (Litzow et al., 2018; Williams and Jackson, 2007; Wolkovich et al., 2014). This possibility is not accounted for in the current study, as our approach assumes stationary relationships among species. We therefore recommend future methodological research to incorporate time-evolving community relationships in our modeling approach in order to examine the potential role of nonstationary community relationships under climate change.

Acknowledgments

NOAA_ERSST_V5 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd/. Groundfish data used here are publicly available from the Alaska Fisheries Science Center at http://www.afsc.noaa.gov/RACE/groundfish/survey_data/data.htm. The R package VAST is publicly available online at https://github.com/James-Thorson/VAST/, and can be used to replicate or update this analysis. We thank S. Zador, M. Scheuerell, and two anonymous reviewers for comments on an earlier draft, and the many scientists who contributed to bottom trawl sampling in the eastern and northern Bering Sea shelf.

Bibliography

Bacheler, N.M., Bailey, K.M., Ciannelli, L., Bartolino, V., Chan, K.-S., 2009. Density-dependent, landscape, and climate effects on spawning distribution of walleye pollock *Theragra chalcogramma*. Mar. Ecol. Prog. Ser. 391, 1–12. https://doi.org/10.3354/meps08259

Baker, M.R., Hollowed, A.B., 2014. Delineating ecological regions in marine systems: Integrating physical structure and community composition to inform spatial management in the eastern Bering Sea. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea III 109, 215–240. https://doi.org/10.1016/j.dsr2.2014.03.001

Bartolino, V., Ciannelli, L., Bacheler, N.M., Chan, K.-S., 2011. Ontogenetic and sex-specific differences in density-dependent habitat selection of a marine fish population. Ecology 92, 189–200. https://doi.org/10.1890/09-1129.1

Bond, N.A., Overland, J.E., Spillane, M., Stabeno, P., 2003. Recent shifts in the state of the North Pacific. Geophys. Res. Lett. 30. https://doi.org/10.1029/2003GL018597

Brown, J., Fiechter, J., 2012. Quantifying eddy–chlorophyll covariability in the Coastal Gulf of Alaska. Dyn. Atmospheres Oceans 55–56, 1–21. https://doi.org/10.1016/j.dynatmoce.2012.04.001

Burrows, M.T., Schoeman, D.S., Buckley, L.B., Moore, P., Poloczanska, E.S., Brander, K.M., Brown, C., Bruno, J.F., Duarte, C.M., Halpern, B.S., Holding, J., Kappel, C.V., Kiessling, W., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Schwing, F.B., Sydeman, W.J., Richardson, A.J., 2011. The Pace of Shifting Climate in Marine and Terrestrial Ecosystems. Science 334, 652–655. https://doi.org/10.1126/science.1210288

Ciannelli, L., Bailey, K.M., 2005. Landscape dynamics and resulting species interactions: the cod-capelin system in the southeastern Bering Sea. Mar. Ecol. Prog. Ser. 291, 227–236. https://doi.org/10.3354/meps291227

Ciannelli, L., Bartolino, V., Chan, K.-S., 2012. Non-additive and non-stationary properties in the spatial distribution of a large marine fish population. Proc. R. Soc. B Biol. Sci. 279, 3635–3642. https://doi.org/10.1098/rspb.2012.0849

Clark, J.S., Gelfand, A.E., Woodall, C.W., Zhu, K., 2014. More than the sum of the parts: forest climate response from joint species distribution models. Ecol. Appl. 24, 990–999. https://doi.org/10.1890/13-1015.1

Colebrook, J.M., 1978. Continuous plankton records-zooplankton and environment, northeast Atlantic and North-Sea, 1948-1975. Oceanol. Acta 1, 9–23.

Cury, P.M., Shin, Y.-J., Planque, B., Durant, J.M., Fromentin, J.-M., Kramer-Schadt, S., Stenseth, N.C., Travers, M., Grimm, V., 2008. Ecosystem oceanography for global change in fisheries. Trends Ecol. Evol. 23, 338–346. https://doi.org/10.1016/j.tree.2008.02.005

Di Lorenzo, E., Schneider, N., Cobb, K.M., Franks, P.J.S., Chhak, K., Miller, A.J., McWilliams, J.C., Bograd, S.J., Arango, H., Curchitser, E., Powell, T.M., Rivière, P., 2008. North Pacific Gyre Oscillation links ocean climate and ecosystem change. Geophys. Res. Lett. 35. https://doi.org/10.1029/2007GL032838

Ding, R., Li, J., Tseng, Y., Sun, C., Guo, Y., 2015. The Victoria mode in the North Pacific linking extratropical sea level pressure variations to ENSO. J. Geophys. Res. Atmospheres 120, 27–45. https://doi.org/10.1002/2014JD022221

Drinkwater, K.F., Beaugrand, G., Kaeriyama, M., Kim, S., Ottersen, G., Perry, R.I., Pörtner, H.-O., Polovina, J.J., Takasuka, A., 2010. On the processes linking climate to ecosystem changes. J. Mar. Syst., Impact of climate variability on marine ecosystems: A comparative approach 79, 374–388. https://doi.org/10.1016/j.jmarsys.2008.12.014

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 1–17. https://doi.org/10.1080/10556788.2011.597854

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., Isaacson, E., 1981. Applications of Estimation Theory to Numerical Weather Prediction, in: Bengtsson, L., Ghil, Michael, Källén, E. (Eds.), Dynamic Meteorology: Data Assimilation Methods, Applied Mathematical Sciences. Springer New York, New York, NY, pp. 139–224. https://doi.org/10.1007/978-1-4612-5970-1_5

Grimmer, M., 1963. The space-filtering of monthly surface temperature anomaly data in terms of pattern, using empirical orthogonal functions. Q. J. R. Meteorol. Soc. 89, 395–408. https://doi.org/10.1002/qj.49708938111

Henson, S.A., Beaulieu, C., Ilyina, T., John, J.G., Long, M., Séférian, R., Tjiputra, J., Sarmiento, J.L., 2017. Rapid emergence of climate change in environmental drivers of marine ecosystems. Nat. Commun. 8, 14682. https://doi.org/10.1038/ncomms14682

Hsieh, C., Reiss, C.S., Hunter, J.R., Beddington, J.R., May, R.M., Sugihara, G., 2006. Fishing elevates variability in the abundance of exploited species. Nature 443, 859. https://doi.org/10.1038/nature05232

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 Nafional Cent. Environ. Informafion.

Hunt, G.L., Coyle, K.O., Eisner, L.B., Farley, E.V., Heintz, R.A., Mueter, F., Napp, J.M., Overland, J.E., Ressler, P.H., Salo, S., Stabeno, P.J., 2011. Climate impacts on eastern Bering Sea foodwebs: a synthesis of new data and an assessment of the Oscillating Control Hypothesis. ICES J. Mar. Sci. 68, 1230–1243. https://doi.org/10.1093/icesjms/fsr036

Kavanaugh, M.T., Hales, B., Saraceno, M., Spitz, Y.H., White, A.E., Letelier, R.M., 2014. Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology. Prog. Oceanogr. 120, 291–304. https://doi.org/10.1016/j.pocean.2013.10.013

Kidson, J.W., 1975a. Eigenvector Analysis of Monthly Mean Surface Data. Mon. Weather Rev. 103, 177–186. https://doi.org/10.1175/1520-0493(1975)103<0177:EAOMMS>2.0.CO;2

Kidson, J.W., 1975b. Tropical Eigenvector Analysis and the Southern Oscillation. Mon. Weather Rev. 103, 187–196. https://doi.org/10.1175/1520-0493(1975)103<0187:TEAATS>2.0.CO;2

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 Res. Part II Top. Stud. Oceanogr. 94, 231–243.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB: Automatic Differentiation and Laplace Approximation. J. Stat. Softw. 70, 1–21. https://doi.org/10.18637/jss.v070.i05

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. Ecol. Lett. 12, 144–154.

Lauth, R.R., Conner, J., 2016. Results of the 2013 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate resources (NOAA Technical Memorandum No. NMFS-AFSC-331). 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. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x

Litzow, M.A., 2017. Indications of hysteresis and early warning signals of reduced community resilience during a Bering Sea cold anomaly. Mar. Ecol. Prog. Ser. 571, 13–28. https://doi.org/10.3354/meps12137

Litzow, M.A., Ciannelli, L., Puerta, P., Wettstein, J.J., Rykaczewski, R.R., Opiekun, M., 2018. Non-stationary climate–salmon relationships in the Gulf of Alaska. Proc R Soc B 285, 20181855. https://doi.org/10.1098/rspb.2018.1855

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*. Bull. Am. Meteorol. Soc. 78, 1069–1080. https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2

Marshall, A.M., Bigg, G.R., Leeuwen, S.M. van, Pinnegar, J.K., Wei, H.-L., Webb, T.J., Blanchard, J.L., 2016. Quantifying heterogeneous responses of fish community size structure using novel combined statistical techniques. Glob. Change Biol. 22, 1755–1768. https://doi.org/10.1111/gcb.13190

McClatchie, S., Gao, J., Drenkard, E.J., Thompson, A.R., Watson, W., Ciannelli, L., Bograd, S.J., Thorson, J.T., 2018. Interannual and Secular Variability of Larvae of Mesopelagic and Forage Fishes in the Southern California Current System. J. Geophys. Res. Oceans 123, 6277–6295. https://doi.org/10.1029/2018JC014011

McCune, B., Root, H.T., 2015. Origin of the dust bunny distribution in ecological community data. Plant Ecol. 216, 645–656. https://doi.org/10.1007/s11258-014-0404-1

Morfin, M., Fromentin, J.-M., Jadaud, A., Bez, N., 2012. Spatio-Temporal Patterns of Key Exploited Marine Species in the Northwestern Mediterranean Sea. PLOS ONE 7, e37907. https://doi.org/10.1371/journal.pone.0037907

Mueter, F.J., Litzow, M.A., 2008. Sea Ice Retreat Alters the Biogeography of the Bering Sea Continental Shelf. Ecol. Appl. 18, 309–320. https://doi.org/10.1890/07-0564.1

Myers, R.A., 1998. When Do Environment–recruitment Correlations Work? Rev. Fish Biol. Fish. 8, 285–305.

Newman, M., Alexander, M.A., Ault, T.R., Cobb, K.M., Deser, C., Di Lorenzo, E., Mantua, N.J., Miller, A.J., Minobe, S., Nakamura, H., Schneider, N., Vimont, D.J., Phillips, A.S., Scott, J.D., Smith, C.A., 2016. The Pacific Decadal Oscillation, Revisited. J. Clim. 29, 4399–4427. https://doi.org/10.1175/JCLI-D-15-0508.1

Nichol, D.G., Kotwicki, S., Wilderbuer, T.K., Lauth, R.R., Ianelli, J.N., 2019. Availability of yellowfin sole Limanda aspera to the eastern Bering Sea trawl survey and its effect on estimates of survey biomass. Fish. Res. 211, 319–330. https://doi.org/10.1016/j.fishres.2018.11.017

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

671 oceanography-dependent vital rates in Bayesian hierarchical models. Can. J. Fish. Aquat.
672 Sci. https://doi.org/10.1139/cjfas-2018-0092
673 Ottersen, G., Kim, S., Huse, G., Polovina, J.J., Stenseth, N.Chr., 2010. Major pathways by which
674 climate may force marine fish populations. J. Mar. Syst., Impact of climate variability on
675 marine ecosystems: A comparative approach 79, 343–360.
676 https://doi.org/10.1016/j.jmarsys.2008.12.013
677 Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., Roslin, T.,
678 Abrego, N., 2017. How to make more out of community data? A conceptual framework
679 and its implementation as models and software. Ecol. Lett. 20, 561–576.
680 https://doi.org/10.1111/ele.12757
681 Planque, B., Arneberg, P., 2018. Principal component analyses for integrated ecosystem
682 assessments may primarily reflect methodological artefacts. ICES J. Mar. Sci. 75, 1021–
683 1028. https://doi.org/10.1093/icesjms/fsx223
684 Puerta, P., Ciannelli, L., Rykaczewski, R.R., Opiekun, M., Litzow, M.A., 2019. Do Gulf of
685 Alaska fish and crustacean populations show synchronous non-stationary responses to
686 climate? Prog. Oceanogr. 175, 161–170. https://doi.org/10.1016/j.pocean.2019.04.002
687 Schirripa, M.J., Goodyear, C.P., Methot, R.M., 2009. Testing different methods of incorporating
688 climate data into the assessment of US West Coast sablefish. ICES J. Mar. Sci. 66, 1605–
689 1613.
690 Schwing, F.B., Mendelssohn, R., Bograd, S.J., Overland, J.E., Wang, M., Ito, S., 2010. Climate
691 change, teleconnection patterns, and regional processes forcing marine populations in the
692 Pacific. J. Mar. Syst., Impact of climate variability on marine ecosystems: A comparative
693 approach 79, 245–257. https://doi.org/10.1016/j.jmarsys.2008.11.027
694 Sigler, M.F., Stabeno, P.J., Eisner, L.B., Napp, J.M., Mueter, F.J., 2014. Spring and fall
695 phytoplankton blooms in a productive subarctic ecosystem, the eastern Bering Sea,
696 during 1995–2011. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding
697 Ecosystem Processes in the Eastern Bering Sea III 109, 71–83.
698 https://doi.org/10.1016/j.dsr2.2013.12.007
699 Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihood in non-
700 Gaussian hierarchical models. Comput. Stat. Data Anal. 51, 699–709.
701 Spencer, P.D., Holsman, K.K., Zador, S., Bond, N.A., Mueter, F.J., Hollowed, A.B., Ianelli, J.N.,
702 2016. Modelling spatially dependent predation mortality of eastern Bering Sea walleye
703 pollock, and its implications for stock dynamics under future climate scenarios. ICES J.
704 Mar. Sci. 73, 1330–1342. https://doi.org/10.1093/icesjms/fsw040
705 Stabeno, P.J., Kachel, N.B., Moore, S.E., Napp, J.M., Sigler, M., Yamaguchi, A., Zerbini, A.N.,
706 2012. Comparison of warm and cold years on the southeastern Bering Sea shelf and some
707 implications for the ecosystem. Deep Sea Res. Part II Top. Stud. Oceanogr.,
708 Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 31–45.
709 https://doi.org/10.1016/j.dsr2.2012.02.020
710 Stenseth, N.Chr., Ottersen, G., Hurrell, J.W., Mysterud, A., Lima, Chan, K., Yoccoz, N.G.,
711 Ådlandsvik, B., 2003. Studying climate effects on ecology through the use of climate
712 indices: the North Atlantic Oscillation, El Niño Southern Oscillation and beyond. Proc.
713 R. Soc. Lond. B Biol. Sci. 270, 2087–2096. https://doi.org/10.1098/rspb.2003.2415
714 Thorson, Jensen, O.P., Hilborn, R., 2015a. Probability of stochastic depletion: an easily
715 interpreted diagnostic for stock assessment modelling and fisheries management. ICES J.
716 Mar. Sci. J. Cons. 72, 428–435. https://doi.org/10.1093/icesjms/fsu127

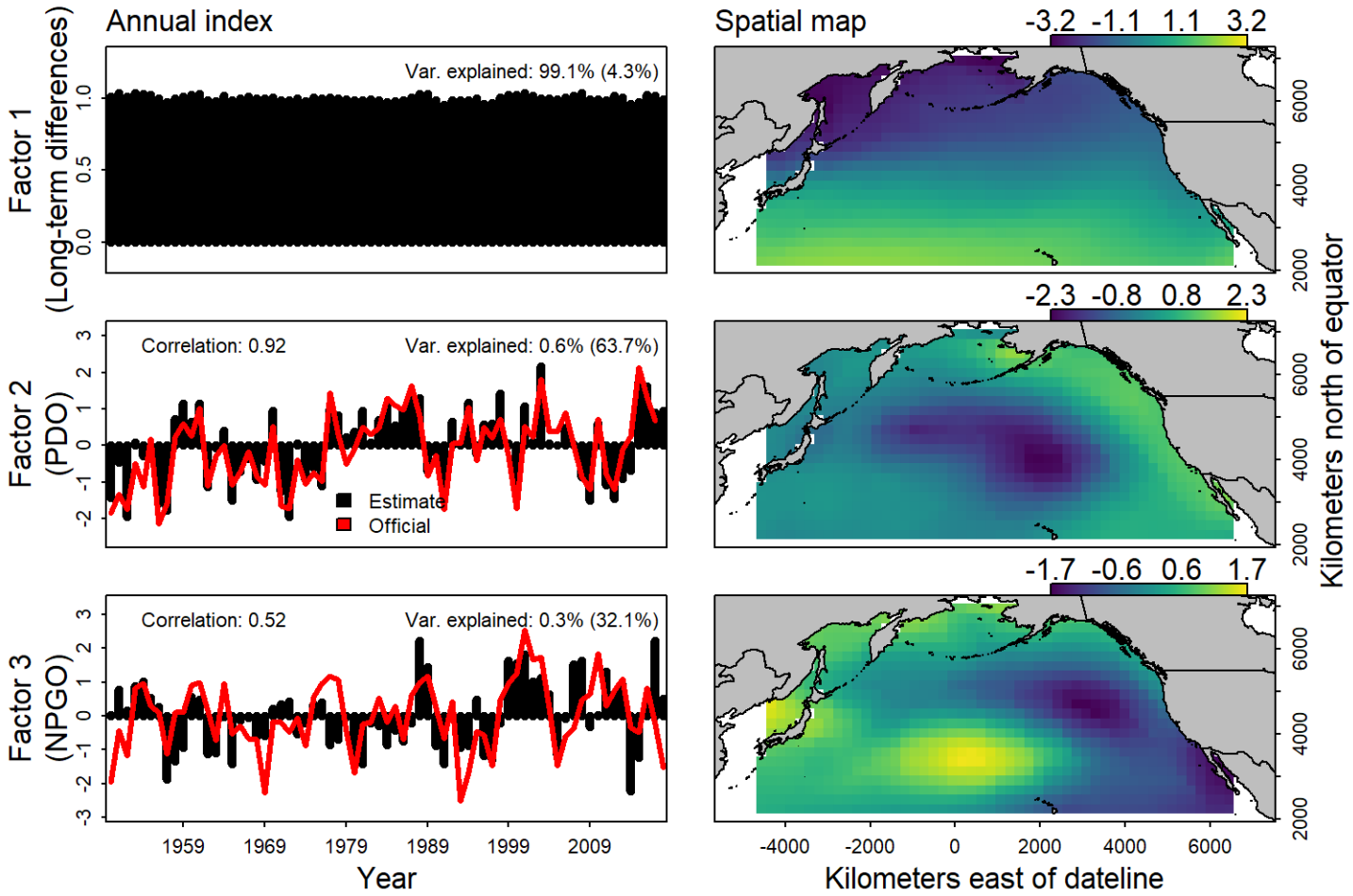
717 Thorson, J.T., 2019a. Forecast skill for predicting distribution shifts: A retrospective experiment
718 for marine fishes in the Eastern Bering Sea. Fish Fish. 20, 159–173.
719 https://doi.org/10.1111/faf.12330
720 Thorson, J.T., 2019b. Guidance for decisions using the Vector Autoregressive Spatio-Temporal
721 (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210,
722 143–161. https://doi.org/10.1016/j.fishres.2018.10.013
723 Thorson, J.T., 2018. Three problems with the conventional delta-model for biomass sampling
724 data, and a computationally efficient alternative. Can. J. Fish. Aquat. Sci. 75, 1369–1382.
725 https://doi.org/10.1139/cjfas-2017-0266
726 Thorson, J.T., In press. Measuring the impact of oceanographic indices on species distribution
727 shifts: The spatially varying effect of cold-pool extent in the eastern Bering Sea. Limnol.
728 Oceanogr. 0. https://doi.org/10.1002/lno.11238
729 Thorson, J.T., Adams, G., Holsman, K., In press. Spatio-temporal models of intermediate
730 complexity for ecosystem assessments: A new tool for spatial fisheries management. Fish
731 Fish. https://doi.org/10.1111/faf.12398
732 Thorson, J.T., Barnett, L.A.K., 2017. Comparing estimates of abundance trends and distribution
733 shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar.
734 Sci. 74, 1311–1321. https://doi.org/10.1093/icesjms/fsw193
735 Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., Winker, H., 2017a. Accounting for
736 spatiotemporal variation and fisher targeting when estimating abundance from
737 multispecies fishery data. Can. J. Fish. Aquat. Sci. 74, 1794–1807.
738 https://doi.org/10.1139/cjfas-2015-0598
739 Thorson, J.T., Ianelli, J.N., Kotwicki, S., 2017b. The relative influence of temperature and size-
740 structure on fish distribution shifts: A case-study on Walleye pollock in the Bering Sea.
741 Fish Fish. 18, 1073–1084. https://doi.org/10.1111/faf.12225
742 Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C., Zipkin, E.F.,
743 2016a. Joint dynamic species distribution models: a tool for community ordination and
744 spatio-temporal monitoring. Glob. Ecol. Biogeogr. 25, 1144–1158.
745 https://doi.org/10.1111/geb.12464
746 Thorson, J.T., Rindorf, A., Gao, J., Hanselman, D.H., Winker, H., 2016b. Density-dependent
747 changes in effective area occupied for sea-bottom-associated marine fishes. Proc R Soc B
748 283, 20161853. https://doi.org/10.1098/rspb.2016.1853
749 Thorson, Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J., Kristensen, K., 2015b. Spatial
750 factor analysis: a new tool for estimating joint species distributions and correlations in
751 species range. Methods Ecol. Evol. 6, 627–637. https://doi.org/10.1111/2041-
752 210X.12359
753 Trenberth, K.E., Fasullo, J.T., Branstator, G., Phillips, A.S., 2014. Seasonal aspects of the recent
754 pause in surface warming. Nat. Clim. Change 4, 911–916.
755 https://doi.org/10.1038/nclimate2341
756 von Szalay, P.G., Somerton, D.A., 2005. The effect of net spread on the capture efficiency of a
757 demersal survey trawl used in the eastern Bering Sea. Fish. Res. 74, 86–95.
758 https://doi.org/10.1016/j.fishres.2005.04.007
759 Walker, GT., 1924. Correlations in seasonal variations of weather. I. A further study of world
760 weather. Mem Indian Meteorol Dep 24, 275–332.

761 Walsh, J.E., Thoman, R.L., Bhatt, U.S., Bieniek, P.A., Brettschneider, B., Brubaker, M.,
762 Danielson, S., Lader, R., Fetterer, F., Holderied, K., 2018. The High Latitude Marine
763 Heat Wave of 2016 and Its Impacts on Alaska. Bull. Am. Meteorol. Soc. 99, S39–S43.
764 Wikle, C.K., Cressie, N., 1999. A dimension-reduced approach to space-time Kalman filtering.
765 Biometrika 86, 815–829. https://doi.org/10.1093/biomet/86.4.815
766 Williams, J.W., Jackson, S.T., 2007. Novel climates, no-analog communities, and ecological
767 surprises. Front. Ecol. Environ. 5, 475–482. https://doi.org/10.1890/070037
768 Wills, R.C., Schneider, T., Wallace, J.M., Battisti, D.S., Hartmann, D.L., 2018. Disentangling
769 Global Warming, Multidecadal Variability, and El Niño in Pacific Temperatures.
770 Geophys. Res. Lett. 45, 2487–2496. https://doi.org/10.1002/2017GL076327
771 Wolkovich, E.M., Cook, B.I., McLauchlan, K.K., Davies, T.J., 2014. Temporal ecology in the
772 Anthropocene. Ecol. Lett. 17, 1365–1379. https://doi.org/10.1111/ele.12353
773 Wyllie-Echeverria, T., Wooster, W.S., 1998. Year-to-year variations in Bering Sea ice cover and
774 some consequences for fish distributions. Fish. Oceanogr. 7, 159–170.
775 https://doi.org/10.1046/j.1365-2419.1998.00058.x
776 Zuur, A.F., Tuck, I.D., Bailey, N., 2003. Dynamic factor analysis to estimate common trends in
777 fisheries time series. Can. J. Fish. Aquat. Sci. 60, 542–552. https://doi.org/10.1139/f03-
778 030
779

780

781

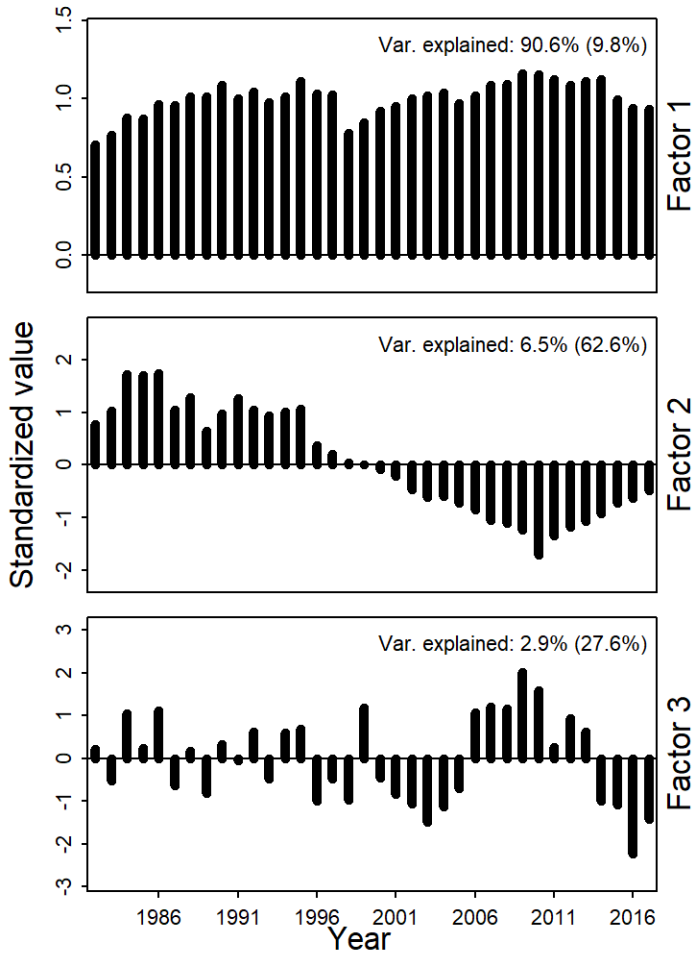
782 Fig. 1 – Visualization of three dominant axes of variability for sea surface temperature in the North Pacific (rows), where each axis
783 includes an index (left columns) showing the magnitude and phase (y-axis) for each year (x-axis), as well as a spatial map (right
784 column) showing the variability in temperature expected during a positive phase, where each map has a spatial variance of 1.0. For
785 each axis we list the proportion of variance explained (top-right of each panel) as well as the proportion of temporal variance
786 explained (calculated after subtracting off the mean across years for each index; listed in parentheses). The second axis is highly
787 correlated with the PDO and the third axis is correlated with the NPGO, so in each case we show the estimated index (black line)
788 relative to the published PDO or NPGO index (red line) and list the correlation between estimated and published indices (top-left of
789 each panel).


Annual index
Spatial map
Factor 1 (Long-term differences)
Var. explained: 99.1% (4.3%)
-3.2 -1.1 1.1 3.2
Factor 2 (PDO)
Correlation: 0.92
Var. explained: 0.6% (63.7%)
Estimate
Official
-2.3 -0.8 0.8 2.3
Factor 3 (NPGO)
Correlation: 0.52
Var. explained: 0.3% (32.1%)
-1.7 -0.6 0.6 1.7
1959 1969 1979 1989 1999 2009
Year
-4000 -2000 0 2000 4000 6000
Kilometers east of dateline
Kilometers north of equator

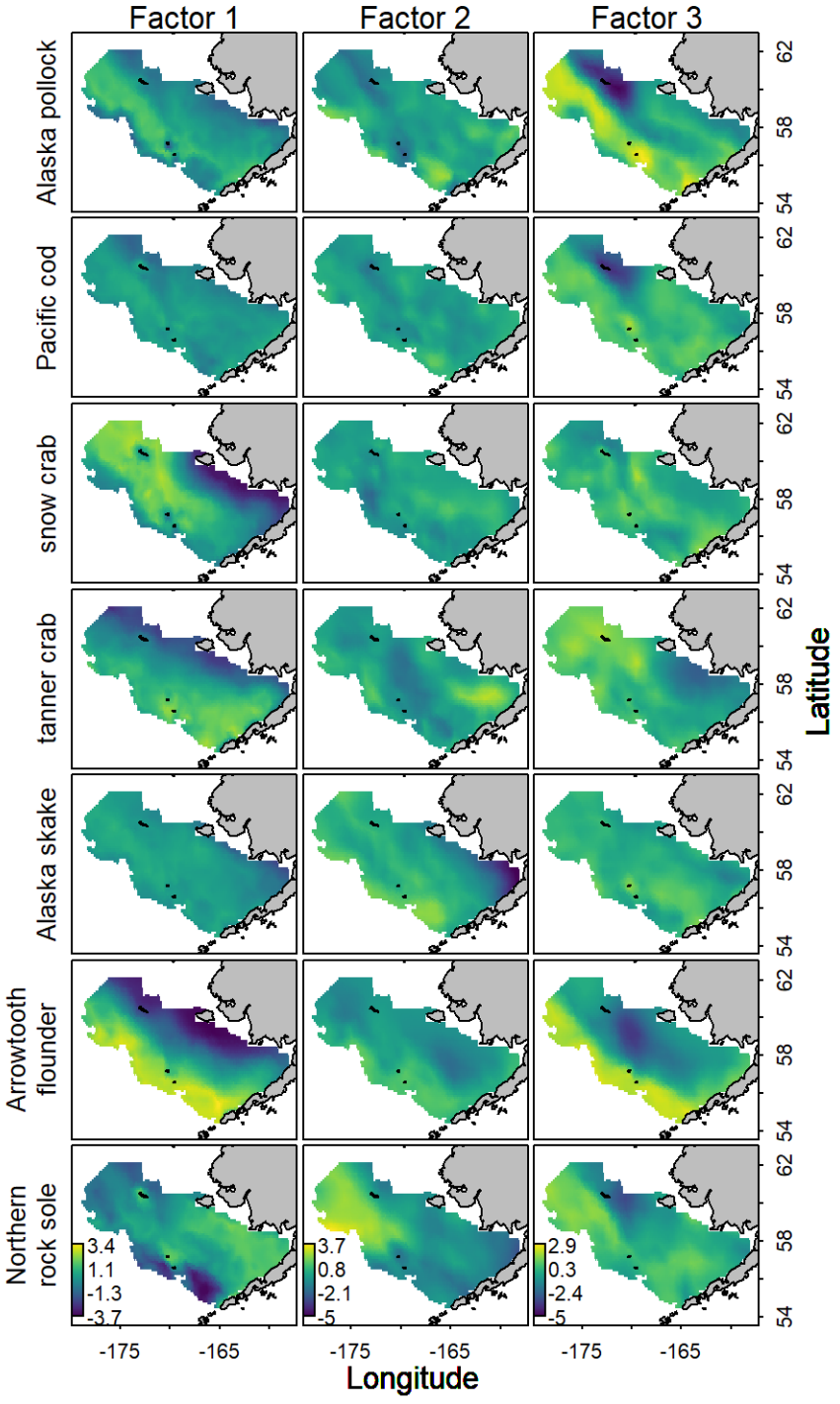
790

Fig. 2 – Visualization of three dominant axes of variability for bottom trawl sampling of biomass for fourteen bottom-associated fish and crab species in the Eastern Bering Sea, where each axis includes an index (Fig. 2a) showing the magnitude and phase (y-axis) for each year (x-axis), as well as a spatial map (Fig. 2b) showing the variability in biomass expected during a positive phase for seven of the fourteen modeled species (rows) and each of the three indices (columns), where each map has a spatial variance of 1.0. For each axis we list the proportion of variance explained (top of Fig. 2a), as well as the proportion of temporal variance explained (calculated after subtracting off the mean across years for each index; parentheses in Fig. 2a). For spatial maps associated with each axis of variability for the remaining seven species, see Fig. S1.

Fig. 2A



802 Fig. 2B:



803

804

Fig. 3 – Visualization of predicted biomass for three of the fourteen modeled species (rows) in six of the thirty-six modeled years. Species are selected from those shown in Fig. 2B, while years are evenly spaced from 1982-2017; for comparison with a model that includes independent spatial and spatio-temporal for each species see Appendix 3 Fig. S2.

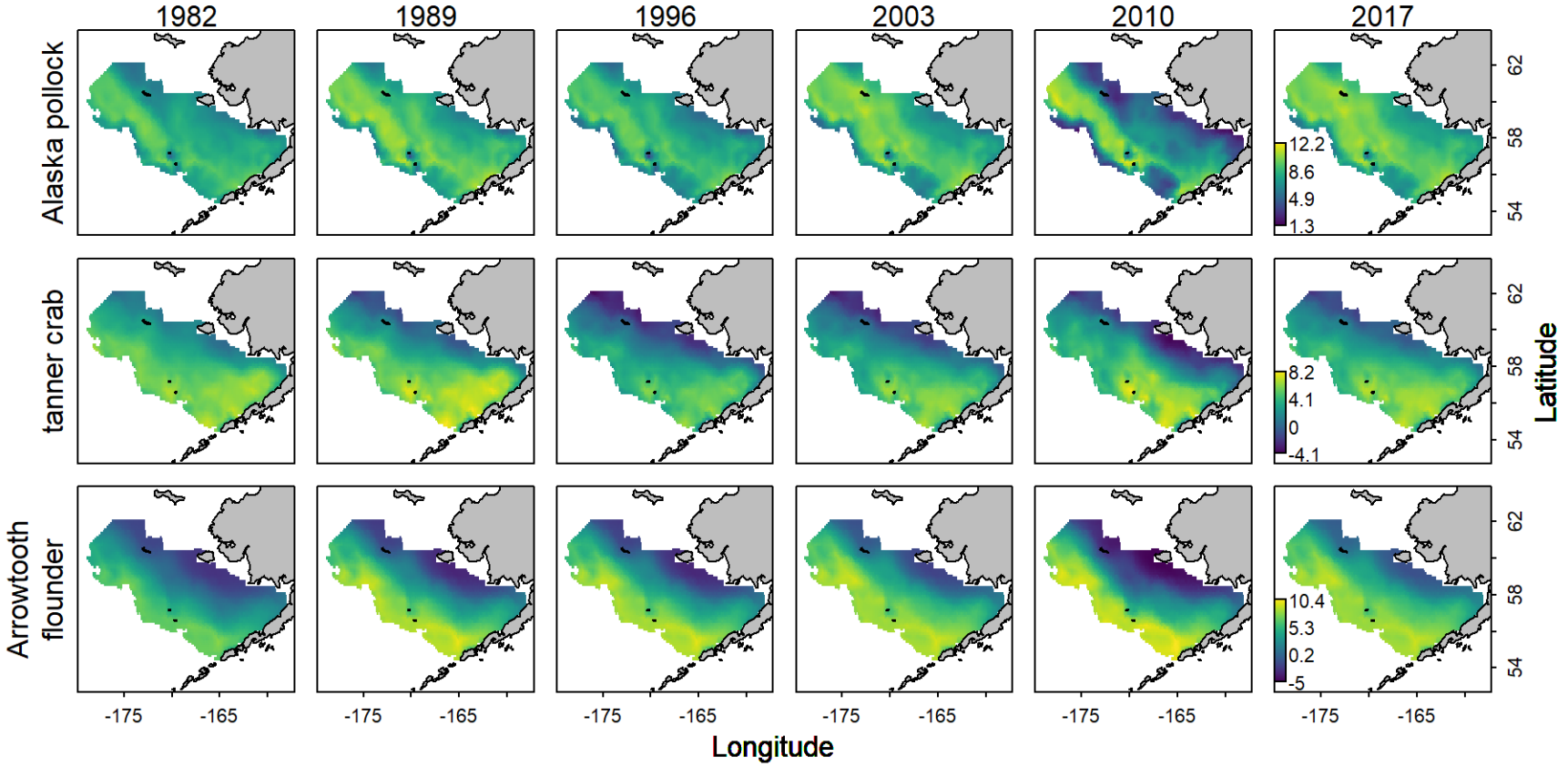


Figure 4 – Comparison of the 3rd estimated index (black bullets) as well ± 1.96 times the estimated standard error in each year (black whiskers), as calculated using Monte Carlo simulation of the estimated loadings matrix $\lambda_f(t)$ and the inverse-Hessian matrix of fixed effects, applied to bottom trawl sampling of biomass for fourteen bottom-associated fishes and crabs in the Eastern Bering Sea (black line; left y-axis label) and the cold pool area (blue line; right y-axis label). The correlation between estimated and published indices is listed (top-left of the plot).