**Forecasting distribution shifts using oceanographic indices: the spatially varying effect of cold-pool extent in the Eastern Bering Sea**

Alternative title: Spatially-varying coefficient models for cold-pool extent explain distribution shifts beyond the effect of local bottom temperature

James T. Thorson

Habitat and Ecosystem Process Research program, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, WA, USA

Journals: Annual Reviews in Marine Science; Progress in Oceanography; Deep-Sea Research II

Keywords: Eastern Bering Sea; Cold pool extent; Spatially varying coefficient; species distribution model; vector-autogressive spatio-temporal; VAST;

**Abstract**

Oceanographers have spent decades developing annual indices that represent basin-scale physical conditions in marine ecosystems. Well-known examples include the Pacific Decadal Oscillation, representing annual variation in the location of warm waters relative to climatologic averages in the North Pacific, or cold-pool extent (CPE), representing the area with cold waters resulting from previous ice melts in the Eastern Bering Sea. However, these indices are rarely included in the species distribution models (SDMs) that are increasingly used to identify and forecast distribution shifts under future climate scenarios. I therefore review three interpretations of spatially varying coefficient (SVC) models, explain how these models can be used to estimate the spatial patterns associated with positive/negative phases of oceanographic indices, and add this option to the multivariate spatio-temporal model *VAST*. I then use a case-study involving bottom trawl data for twenty fish and decapod species in the Eastern Bering Sea 1982-2017 to answer: does an SVC model for cold-pool area explain a substantial portion of variation in spatial distribution for species in this region? and (2) does a spatially varying effect of the cold-pool remain useful and significant even when local temperature is also included as a covariate? Results show that CPE and local bottom temperature are selected by AIC for 13 of 17 species, explain nearly 20% of spatio-temporal variation on average and are parsimonious, and CPE does explain variation in excess of local temperature alone. I therefore conclude that spatially varying coefficient models are a useful way to assimilate oceanographic indices within SDM models, and hypothesize that these will be useful to account for decadal-scale variability within multi-decadal forecasts of distribution shift.

**Introduction**

Ecosystem-based management (EBM) involves regulating multiple ocean impacts including harvest, tourism, and energy development while accounting for species interactions, ecosystem drivers, and socio-economic linkages. EBM has evolved in tandem with new tools for understanding the multiple human impacts and outcomes resulting from ocean management. Examples of new tools include: techniques to include annual oceanographic conditions in the stock-assessment models that are used to define annual catch limits (Schirripa et al., 2009); ecosystem models that include mechanistic detail regarding terrestrial and physical drivers (Fulton et al., 2011); and spatial models that are used to estimate and validate maps of fish habitats (Rooper et al., 2016). Continued improvement in these tools is likely to support ongoing developments in ocean management and governance, e.g., the Bering Sea Fisheries Ecosystem Plan which now includes explicit linkages to climate-linked ecosystem models (citation).

One topic of growing importance in ocean governance is climate-driven shifts in species distribution (Pinsky et al., 2018). Distribution shifts are increasingly identified from multiple data sets using species distribution/density models (SDMs) fitted to occurrence, count, or biomass-sampling data (e.g., Dolder et al., 2018). SDMs have been used extensively in ecology, oceanography, and fisheries science to describe the spatial distribution and ecological niche of marine and terrestrial species worldwide. Interest in SDMs has increased as researchers have sought to train them using historical and contemporary data and then forecast changes in spatial distribution under alternative climate scenarios (Araújo and New, 2007). SDMs have been built to predict local density using local environmental conditions and/or via the interaction of annual covariates and spatial coordinates, and size- or age-structured SDMs have also been used to account for size-based processes affecting distribution shifts (Kristensen et al., 2014; Thorson et al., 2015; Kai et al., 2017).

However, studies examining forecasted distribution shifts that are compared with subsequent observations have shown that forecast skill is sometimes poor using the current generation of SDMs that are fitted to localized effects of environmental conditions (Thorson, 2019a). For example, ontogenetic habitat preferences combined with changes in size/age-structure has been hypothesized to drive observed distribution shifts (Barbeaux and Hollowed, 2018), but this has not been successful at explaining a large portion of historical distribution shifts for *Gadus chalcogrammus* (Thorson et al., 2017) or *Paralichthys dentatus* (Perretti and Thorson, In review). This poor skill would arise if species distributions are by more than local environmental conditions, e.g., due to impacts of local predator densities, lagged environmental effects, or geographically distant environmental conditions that affect the preference of mobile species for local conditions. There is therefore an ongoing need to identify improved techniques to forecast distribution shifts.

In particular, multi-decadal shifts in distribution are typically forecasted by fitting to local environmental conditions (Pinsky et al., 2018), some assumed relationship with local conditions (Cheung et al., 2008), or theoretical predictions of metabolic constraints (Teal et al., 2018). Although these local environmental conditions may capture long-term changes in the fundamental niche, it is likely that decadal-scale oscillations in oceanographic conditions will also contribute to decadal-scale fluctuations around long-term trends. Physical oceanographers have spent decades developing oceanographic indices that are an integrated and high-level representation of interannual variation in ocean conditions (Grimmer, 1963; Kidson, 1975), and fisheries scientists have shown that these oceanographic indices can be informative about spatial patterns in fish productivity (O’Leary et al., 2018). For example, the Pacific Decadal Oscillation (PDO) is an annual index of the location of elevated ocean temperatures in the North Pacific relative to their climatological average, and the PDO was originally identified to correlate with oscillating productivity of salmon stocks between the US West Coast and Alaska (Mantua and Hare, 2002).

Despite the well-documented role of oceanographic indices in driving changes in fish productivity and distribution, there is surprisingly little research regarding their potential role in species distribution models used to forecast marine distribution shifts. As I will show, this may arise due to the way in which oceanographic indices would typically be included in regression models such as SDM, wherein an annual index can only impact spatial distribution though its interaction with a spatially referenced variable. We therefore describe a novel approach to including oceanographic indices via a “spatially varying coefficient” (SVC) model, where the map associated with positive/negative values of the annual index are freely estimated. We first discuss several potential interpretations of this SVC model and modify an existing R package *VAST* for multivariate spatio-temporal models (Thorson and Barnett, 2017) to include SVC models. We then explore SVC models using a case-study involving twenty fish and decapod species in the Eastern Bering Sea, where spatial dynamics are widely hypothesized to be linked to the spatial extent of cold waters resulting from ice melt in previous years. We specifically use this case study to answer two questions: (1) does an SVC model for cold-pool area explain a substantial portion of variation in spatial distribution for species in this region? and (2) does a spatially varying impact of the cold-pool remain useful and significant even when also including local temperature as a covariate? Results confirm that a spatially varying coefficient model for cold-pool extent is informative even given inclusion of local bottom temperature, and that the two variables can explain nearly 20% of spatio-temporal variability across species.

**Methods**

I seek to determine whether spatially varying responses to annual oceanographic indices can (1) improve model parsimony, reduce unexplained variation, and decrease forecast errors relative to a model without covariates and (2) whether spatially varying coefficients are still estimable and parsimonious even when including local environmental conditions in addition to annual oceanographic indices. To do so, I first introduce spatially varying coefficient models, their interpretation, and their estimation. We then describe a case study involving twenty fish and decapod species in the Eastern Bering Sea that is used to address our two questions.

**Spatially-varying coefficient models**

One alternative to using local covariate effects is “spatially varying coefficient” (SVC) models. These models have been discussed extensively in statistical applications (Gelfand et al., 2004; Finley, 2011) and have seen some limited use in fisheries (Bacheler et al., 2009; Bartolino et al., 2011), e.g., to identify seasonal shifts in pollock distribution in the Eastern Bering Sea (Bacheler et al., 2012). As a simple example, SVC models typically predict some response variable as a linear function of a predictor where the parameter representing the linear response of to changes in varies as a function of space:

|  |  |
| --- | --- |
|  | (1) |

where is a global intercept, is the slope at spatial location , and represents residual errors. Although can be approximated as a parametric function (i.e., as a function of some other covariate), a general treatment allows slope to vary randomly across space, , where is the average slope across space and is a zero-mean function representing spatial variation in . In the following, I specify that follows a zero-mean Gaussian Markov random field, , where is the vector of at some set of locations and is a sparse approximation to a spatial correlation function. This specification then allows for easy integration into existing Vector Autoregressive Spatio-Temporal modelling software, available as the R package *VAST* (Thorson, 2019b). Although I focus on implementing this approach using *VAST*, the following could also be done in any other software package that can estimate the linear interaction of a continuous variable and a spatial random field or penalized spline (e.g., package *mgcv*, Wood (2006)).

**Interpretation of spatially varying coefficients in spatio-temporal models**

In the following, I will use spatially varying coefficients to try and explain spatio-temporal variation as resulting from an annual time-series. Annual time-series are widely used to summarize oceanographic conditions in a given marine ecosystem, e.g., where the location of the North Atlantic Oscillation (NAO) is correlated with the Gulf Stream and therefore affects ocean productivity in the Northwest Atlantic shelf (O’Leary et al., 2018). I therefore describe a spatio-temporal model for a response :

|  |  |
| --- | --- |
|  | (2A) |

where is an intercept that varies for every modeled time period (e.g., among years), is persistent spatial variation, and is spatial variation that changes over time (termed “spatio-temporal” variation). Spatial shifts in are caused entirely by , such that the spatial distribution of could be forecasted exactly given perfect information regarding . In the following, I provide three alternative interpretations of spatially varying coefficients:

*Regression of spatio-temporal variation on an annual index*

Analysts include covariates for many different reasons, although I focus in the following on efforts to explain spatio-temporal variation . When covariate represents an annual oceanographic index, it will have the same value for all locations in a given year, such that the product will explain variation in intercepts but have no impact on spatial distribution. We therefore introduce a spatially varying coefficient for annual time series :

|  |  |
| --- | --- |
|  | (2B) |

In many instances, intercepts are estimated as fixed effects and in this instance, is confounded with such that I specify to allow intercepts to be identified. In other instances when has some hierarchical structure across time (e.g., follows an autoregressive process) then and can be separately estimated.

Defining , it is apparent that the spatially varying coefficient defines spatio-temporal variation for all locations and times. If and in Eq. 2A are typically similar for any pair of years and that have similar values and for annual index , then estimating in Eq. 2B will explain the variance that would otherwise be attributed to in Eq. 2A. In this sense, then, defines a linear model for explaining spatio-temporal residuals , and including a spatially varying effect of will reduce the variance of residual spatio-temporal variation .

*Random slope models*

Alternatively, we can interpret spatially varying coefficients as a “random slope” model. Gelman and Hill (2007) define “random intercept” models as linear models that include random variation in the intercept, i.e., by interpreting in Eq. 1 as a spatially varying slope. In contrast, “random slope” models include variation in the slope coefficient, i.e., by treating as a spatially varying slope. Therefore, defining spatially varying coefficients allows an analyst to estimate both a random-intercept and random-slope within SDMs.

*Approximation to functional analysis*

This spatially varying coefficient model could be generalized by estimating a non-parametric “functional” that takes as input a spatial variable in each year and provides as output , representing the impact of this variable at all locations on population density. Using this functional, impact for a given location depends upon the value of at all locations, including geographically distant locations, and not simply upon at that location. However, functional is many-to-many and methods exist to simplify . For example, autoencoders apply a function to compress variable to a smaller set of features , and then a function to project those features on their impact, and is it immediately obvious that functional . Given this expanded notation, our spatially varying coefficient represents the projection function , while an annual index (e.g., the Pacific Decadal Oscillation, PDO) is typically calculated as a compression of some spatially varying measurements (e.g., surface temperature in the North Pacific for the PDO). For example, the cold-pool extent in the Eastern Bering Sea is defined as the total area with bottom temperatures less than or equal to C, and can be calculated from inputted bottom temperatures by compression function where is an indicator function that equals one if and zero otherwise.

**Including spatially varying coefficients in VAST**

In the following, I will analyze biomass-sampling data using a delta model in the R package *VAST*:

where is predicted encounter probability for observation , is predicted biomass given an encounter, is a probability density function for positive catches, and is the residual (“measurement”) variance for samples for category (where categories could represent, e.g., multiple species within a multispecies model). We specifically use a Poisson-link delta model (Thorson, 2017) that includes two linear predictors:

where is predicted numbers-density, is the area sampled for observation , and is predicted biomass-per-individual. and are each specified as log-linked linear models:

where a similar log-linked linear predictor is defined for . For completeness in describing the new SVC feature, I introduce notation for multivariate spatio-temporal models for each of categories where VAST is capable of estimating a rank-reduced covariance among categories via loadings matrices and for multiple “factors” , and where and represent spatial and spatio-temporal variation in at locations and times . In univariate models, and then become scalars representing the standard deviation of spatial or spatio-temporal variation.

VAST now includes term for each of covariates, while previous versions of VAST had instead specified the reduced term . This extended capability for density covariates includes , which is specified as a zero-mean Gaussian Markov random field with unit variance for each category and covariate . When estimating a spatially varying coefficient, VAST estimates its variance as a fixed effect and estimating results in the reduced, linear effect of that covariate. Users must specify that for any covariate that is identical across space whenever intercepts are estimated as fixed effects, and we do this when fitting to cold-pool extent; VAST includes user settings to specify a model where .

**Case-study application**

We explore the utility of this new spatially varying coefficient model using bottom-trawl survey data for fish and decapod species in the Eastern Bering Sea. Previous oceanographic research has shown that the productivity and spatial structure of this marine ecosystem is strongly tied to both the spatial extent of sea ice and the seasonal timing of ice melt. Specifically, sea ice provides a spatial structure for the growth of highly productive phytoplankton, affects the timing of vertical mixing and resulting plankton community structure (Stabeno et al., 2012). Areas with sea ice are also associated with colder bottom temperatures, affecting distribution for marine species that either prefer or avoid near-zero water temperatures (Hunt et al., 2011). We follow past studies in measuring the cold-pool as the area with bottom temperatures at or below C, and explore using this cold-pool area as an annual oceanographic index to describe spatio-temporal variation in distribution.

Specifically, I seek to answer: (1) does a spatially varying coefficient model for cold-pool area explain a substantial portion of variation in spatial distribution for species in this region? and (2) does a spatially varying impact of the cold-pool remain useful and significant even when also including local temperature as a covariate? To answer these questions, I use data for the seventeen numerically dominant, bottom-associated fish and decapod species in the Eastern Bering Sea slope bottom trawl survey conducted annually from 1982 onward using 83-112 bottom trawl gear by the Alaska Fisheries Science Center (Lauth and Conner, 2016). This survey uses a fixed station design involving nearly 370 stations on a 20 km by 20 km grid (including some areas with more dense sampling near significant islands), and data are publicly available online (<http://www.afsc.noaa.gov/RACE/groundfish/survey_data/data.htm>). For each species, I fit eight models, formed as a 4 by 2 factorial cross of four model structures and two data sets. The four model structures include:

1. *No covariates*: A model with spatial and spatio-temporal variation for both numbers density and average weight but no habitat covariates. We model spatio-temporal variation as following a 1st order autoregressive process, such that hotspots in density are predicted to persist for subsequent years, where the degree of autoregression is estimated from available data.
2. *Local temperature*: Identical to *No covariates* except also including a quadratic effect of local bottom temperature for each location and year , representing a dome-shaped response of local density to local temperatures.
3. *Cold*-*pool*: Identical to *No covariates* except also including a spatially varying coefficient linking numbers density and average weight to cold-pool extent .
4. *Both*: Identical to *No covariates* except including both a quadratic effect of local temperature and a spatially varying effect of cold-pool extent.

We fit each of these four models to two different data sets for each species:

1. *Full data*: We fit to all bottom-trawl data 1982-2017 available online in Jan. 2018. Given that I am predicting density for this same set of years 1982-2017, I do not need to impose any restrictions of model intercepts, and , and treat them as fixed effects.
2. *Reduced data:* We fit to data 1982-2014 and forecast future distribution shifts 2015-2017. Given that I have no data to inform intercept estimates for forecast years, I impose a hyperdistribution on model intercepts, which I specify as following a random-walk process where the variance of this random walk is estimated from available data.

We fit all models using R package *VAST* (Thorson, 2019b; Thorson and Barnett, 2017) using Microsoft Open R (R Core Team, 2017). *VAST* estimates fixed effects using *TMB* (Kristensen et al., 2016) and also applies a stochastic partial differential equation approximation to a Matern spatial correlation function (Lindgren et al., 2011) to reduce computational requirements. We use 200 “knots” while confirming that results are qualitatively similar even when changing this number.

**Results**

Plotting the spatially varying effect of cold-pool extent ( and ) shows how distribution is expected to differ in years with high cold-pool extent relative to years with average conditions. Visualizing this effect for four biologically diverse species shows the varied impacts of the cold-pool on species distribution in the Eastern Bering Sea (Fig. 3). *Gadus chalcogrammus* shows increased densities in the southern outer domain near the Aleutian Islands during years with high cold-pool extent, and this affects expected biomass per individual () more strongly than numbers density (). *Hippoglossus stenolepsis* shows elevated numbers-density (and therefore encounter probability) in the middle domain during years with high cold-pool extent, but also elevated biomass-per-encounter in nearshore waters. By contrast, *Atherestes stomias* shows elevated numbers-density and biomass-per-encounter in the outer domain during years with large cold-pool extent, and *Chionectes bairdi* shows relatively little response in numbers-density (and therefore encounter probability) in years with a large cold-pool (i.e., the color-scale in the bottom-left panel of Fig. 3 is very small), while biomass-per-encounter are elevated in the southern middle and outer domain during these years. These responses are just a few of the varied impacts occurring across all seventeen species analyzed here (Fig. S1).

Cold-pool extent explains a portion of variation in log-density for species in the Eastern Bering Sea even when including a quadratic effect of bottom temperature (Fig. 4), although the standard deviation of variation explained by bottom temperature is nearly twice as great on average (0.37/0.27; Fig. 4 2st row) as the spatially varying link to the cold-pool (0.16/0.14; Fig. 4 1st row). Including both bottom temperature and cold-pool extent explains 17% of the residual spatio-temporal variation for numbers density (Fig. 4 bottom-left panel) and 19% of residual spatio-temporal variation for average biomass (Fig. 4 bottom-right panel). Local temperature in isolation explains only 12% and 11% of residual spatio-temporal variation in numbers-density and average biomass, respectively, and both cold-pool extent and bottom temperature contribute nearly equally to the overall reduction in variance (i.e., as shown by the similar reduction for “Temp” and “ColdPool” models in bottom row of Fig. 4).

Relatedly, the Akaike Information Criterion (AIC) indicates that including both local bottom temperature and annual cold-pool extent is parsimonious for thirteen of seventeen species (Table 1), and selects local temperature or cold-pool extent individually for all remaining species. Comparing a model with cold-pool extent to the AIC selected model (i.e., comparing “ColdPool” with “Temp” for *Limanda aspera, Hyas lyratus, or Lycodes brevipes*) shows that including the cold-pool degrades model parsimony very little even when it’s not selected, and exploration shows that the estimate of and is approaching zero in these instances.

Finally, examining forecasts of northward center-of-gravity when fitting to biomass-sampling data for 1982-2015 and forecasting distribution shifts in 2015-2018 (Fig. 5) shows instances where incorporating temperature and cold-pool extent improves forecasts of distribution shift from 2015-2017. For example, *Lycodes palearis* shows a southward shift in 2015 followed by a rapid shift 100 km north 2016-2017. This northward shift is forecasted by all models because the species has a more southward distribution than its long-term average and the autocorrelation in spatio-temporal variation causes it to revert to its long-term average in all models. However, models including local temperature forecast a substantial northward shift 2015-2016, in accordance with subsequent measurements. Local temperature is similarly helpful in forecasting the northward shift for *Mallotus villosus*, *Hippoglosus stenolepsis* and *Hyas coarctatus* in 2016. Meanwhile, cold-pool extent is helpful in improving forecasts of northward distribution in 2015-2016 for *Gadus macrocephalus* relative to models with only local temperature effects. Comparing the error in forecasted center-of-gravity among models (Fig. 6) shows that including temperature and cold-pool extent decreases the median absolute error by 25% (from 15.7 to 11.8 km) for one-year forecasts, 37% for two-year forecasts (from 30.5 to 19.3 km) and 5% for three-year forecasts (from 35.6 to 33.9 km) relative to a model with neither covariate, and similarly decreases bias in forecasts for each year. The model without covariates itself has similar or lower errors than the persistence forecast, due to its inclusion of density dependence as represented by the autoregressive structure on density hotspots.

**Discussion**

In this study, I have shown that spatially varying coefficient models can be used to predict and forecast changes in spatial distribution using annual oceanographic indices. Using data for seventeen bottom-associated fish and crab species in the Eastern Bering Sea, I have shown that a dome-shaped response to local bottom temperature explains 11-12% of spatio-temporal variation in two predictors of density, while local temperature and the additional, spatially varying impact of cold-pool extent explains 17-19% of variation in these two predictors. In addition, both local temperature and cold-pool effects are identified as parsimonious by AIC for thirteen of the seventeen species, and including both effects generally improves 1-, 2-, or 3-year forecasts of distribution over a model without covariates or a persistence forecast. I therefore conclude that spatially varying coefficients are a useful tool in the toolbox of methods available for explaining and forecasting distribution within species distribution models, and are particularly valuable as a way to integrate regional oceanographic indicators.

Given the benefit of including spatially varying effects of oceanographic indices, I also highlight two paradoxes of our results. Specifically, cold-pool area is calculated from local measurements of bottom temperature and represents a highly compressed representation which “loses” information about annual differences in the spatial distribution of cold habitats. However, our analysis shows that cold-pool area contains additional information about species distribution beyond the information contained in local temperatures. How do fishes “know” about temperatures occurring in habitats outside the area being sampled? And how does compressing temperature measurements result in “more information?” We address these two questions below.

Regarding how fishes “know” about temperatures occurring in habitats outside the area being sampled: the fish and decapods analyzed here are highly mobile and select their habitats after having “sampled” (migrated through or foraged within) other habitats available regionally. This answer is obvious biologically, and community ecology acknowledges many non-local and regional mechanisms are needed to explain species distribution and density (HilleRisLambers et al., 2013; Heino et al., 2017). However, SDMs have typically focused on including local habitat measurements within regression models for predicting local densities, and these non-local mechanisms are rarely included explicitly in SDMs. Alternatively, some SDMs apply a kernel smoother to local covariates prior to inclusion (Chandler and Hepinstall-Cymerman, 2016; Frishkoff et al., In press); this smoother captures the response to a spatial average of nearby habitats, but does not allow for complicated dependencies on spatially distant habitats like the spatially varying coefficient model explored here. I therefore argue that spatially varying coefficients for annual oceanographic indices represent a useful and flexible way to represent habitat selection as animals respond to information about resource availability in years with different oceanographic conditions. For example, pollock in the Bering Sea have been shown to distribute along the sea-ice break during spring (De Robertis and Cokelet, 2012), and the location of the ice-break is driven by several region-scale and lagged processes that are not necessarily related to local bottom temperatures. In these cases, a regional covariate (e.g., sea-ice extent) represents the integrated effect of lagged and geographically distant habitat conditions that affect fish decisions regarding habitat selection.

Regarding how compressing temperature measurements result in more information: the spatially varying effect of cold-pool area upon numbers-density and average-weight (Fig. 3) differs greatly among species in a way that cannot be explained by the local impact of bottom temperature. For example, *Gadus chalcogrammus* has reduced densities in the northern boundary in years with high cold-pool extent, and this clearly mirrors the areas where bottom temperatures are strongly impacted by cold-pool extent. However, *Clupea pallasii* and *Hippoglossus stenolepsis* show contrasting effects of cold-pool extent and neither exactly coincides with areas that show greatest changes in local temperature. We therefore see a spatially varying response to cold-pool area as an approximation to regional temperature impacts that vary among species. In particular, compressing temperature to calculate cold-pool area and then using a species-specific projection of this index on distribution approximates a flexible “functional” translating both local and non-local temperatures to densities maps (see *Interpretation of spatially varying coefficients in spatio-temporal models* for more details). We therefore encourage ongoing research to allow spatio-temporal models to simultaneously estimate a generalized “compression” function calculating regional indices from local environmental conditions, in addition to the nonparametric “projection” function that spatially varying coefficient models already involve. This could perhaps be accomplished by simultaneously estimating parameters for the “empirical orthogonal function” models that are typically used to calculate oceanographic indices and the SDMs that is used to estimate species distributions.

In particular, a spatially-varying coefficient for regional oceanographic conditions allows analysts to assimilate both fine-scale and regional covariates in a single modelling framework, rather than building separate models for each resolution (e.g., Smart et al., 2012). Alternatively, previous studies have included regional covariates while estimating a separate spatial distribution for “warm” and “cool” years in the Eastern Bering Sea (e.g., Hollowed et al., 2012), while the SVC model avoids classifying regional conditions into a factor prior to analysis. In the Eastern Bering Sea, cold-pool extent is associated with many processes that might be associated mechanistically with habitat selection and resulting fish distribution including: plankton prey densities (Stabeno et al., 2012), dominant modes of physical transport (Zhang et al., 2012), and partitioning of available primary productivity between pelagic and demersal communities (Hunt et al., 2011). The SVC allows these many upstream and downstream processes affecting distribution to be summarized in an annual index that is used to hindcast and forecast spatial patterns in community structure.

Most importantly, I see spatially varying coefficients for oceanographic indices as an important programmatic tool to translate climatological research from physical oceanography into the distribution models that are used for marine spatial planning (Rassweiler et al., 2014), essential fish habitat (Rooney et al., 2018), and ecosystem based management (Link and Browman, 2014). Methods exist for translating indices like the Pacific Decadal Oscillation into stock assessment models used for defining harvest limits (Schirripa et al., 2009), and these methods are increasingly used to evaluate ecological claims regarding bottom-up limits for fisheries productivity (O’Leary et al., 2018). However, oceanographic indices are less commonly used in SDM forecasts of distribution shift. We hope that demonstrating the benefit of oceanographic indices in species distribution models will continue to “build a bridge” between fisheries oceanography and ecosystem-based management.

**Acknowledgements**

I thank M. Litzow, L. Rogers, and L. Ciannelli for discussions that prompted me to investigate spatially varying coefficient models, as well as sparked my interest in combining physical-oceanography with species distribution models. I also thank C. Rooper and D. McGowan for comments on an earlier draft.

**Works cited**

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47. https://doi.org/10.1016/j.tree.2006.09.010

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

Bacheler, N.M., Ciannelli, L., Bailey, K.M., Bartolino, V., 2012. Do walleye pollock exhibit flexibility in where or when they spawn based on variability in water temperature? Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 208–216. https://doi.org/10.1016/j.dsr2.2012.02.001

Barbeaux, S.J., Hollowed, A.B., 2018. Ontogeny matters: Climate variability and effects on fish distribution in the eastern Bering Sea. Fish. Oceanogr. 27, 1–15. https://doi.org/10.1111/fog.12229

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

Chandler, R., Hepinstall-Cymerman, J., 2016. Estimating the spatial scales of landscape effects on abundance. Landsc. Ecol. 31, 1383–1394. https://doi.org/10.1007/s10980-016-0380-z

Cheung, W.W.L., Lam, V.W.Y., Pauly, D., 2008. Modelling present and climate-shifted distribution of marine fishes and invertebrates.

De Robertis, A., Cokelet, E.D., 2012. Distribution of fish and macrozooplankton in ice-covered and open-water areas of the eastern Bering Sea. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 217–229. https://doi.org/10.1016/j.dsr2.2012.02.005

Dolder, P.J., Thorson, J.T., Minto, C., 2018. Spatial separation of catches in highly mixed fisheries. Sci. Rep. 8, 13886. https://doi.org/10.1038/s41598-018-31881-w

Finley, A.O., 2011. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. Methods Ecol. Evol. 2, 143–154. https://doi.org/10.1111/j.2041-210X.2010.00060.x

Frishkoff, L.O., Mahler, D.L., Fortin, M.-J., In press. Integrating Over Uncertainty In Spatial Scale Of Response Within Multispecies Occupancy Models Yields More Accurate Assessments Of Community Composition. Ecography.

Fulton, E.A., Link, J.S., Kaplan, I.C., Savina-Rolland, M., Johnson, P., Ainsworth, C., Horne, P., Gorton, R., Gamble, R.J., Smith, A.D.M., Smith, D.C., 2011. Lessons in modelling and management of marine ecosystems: the Atlantis experience. Fish Fish. 12, 171–188. https://doi.org/10.1111/j.1467-2979.2011.00412.x

Gelfand, A.E., Schmidt, A.M., Banerjee, S., Sirmans, C.F., 2004. Nonstationary multivariate process modeling through spatially varying coregionalization. Test 13, 263–312. https://doi.org/10.1007/BF02595775

Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, UK.

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

Heino, J., Soininen, J., Alahuhta, J., Lappalainen, J., Virtanen, R., 2017. Metacommunity ecology meets biogeography: effects of geographical region, spatial dynamics and environmental filtering on community structure in aquatic organisms. Oecologia 183, 121–137. https://doi.org/10.1007/s00442-016-3750-y

HilleRisLambers, J., Harsch, M.A., Ettinger, A.K., Ford, K.R., Theobald, E.J., 2013. How will biotic interactions influence climate change–induced range shifts? Ann. N. Y. Acad. Sci. 1297, 112–125.

Hollowed, A.B., Barbeaux, S.J., Cokelet, E.D., Farley, E., Kotwicki, S., Ressler, P.H., Spital, C., Wilson, C.D., 2012. Effects of climate variations on pelagic ocean habitats and their role in structuring forage fish distributions in the Bering Sea. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 230–250. https://doi.org/10.1016/j.dsr2.2012.02.008

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

Kai, M., Thorson, J.T., Piner, K.R., Maunder, M.N., 2017. Spatio-temporal variation in size-structured populations using fishery data: an application to shortfin mako (Isurus oxyrinchus) in the Pacific Ocean. Can. J. Fish. Aquat. Sci. https://doi.org/10.1139/cjfas-2016-0327

Kidson, J.W., 1975. 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

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

Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E., 2014. Estimating spatio-temporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. 71, 326–336. https://doi.org/10.1139/cjfas-2013-0151

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

Link, J.S., Browman, H.I., 2014. Integrating what? Levels of marine ecosystem-based assessment and management. ICES J. Mar. Sci. 71, 1170–1173. https://doi.org/10.1093/icesjms/fsu026

Mantua, N.J., Hare, S.R., 2002. The Pacific decadal oscillation. J. Oceanogr. 58, 35–44.

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. Can. J. Fish. Aquat. Sci. https://doi.org/10.1139/cjfas-2018-0092

Perretti, C.T., Thorson, J.T., In review. Spatio-temporal dynamics of summer flounder (Paralichthys dentatus) on the Northeast U.S. Shelf.

Pinsky, M.L., Reygondeau, G., Caddell, R., Palacios-Abrantes, J., Spijkers, J., Cheung, W.W.L., 2018. Preparing ocean governance for species on the move. Science 360, 1189–1191. https://doi.org/10.1126/science.aat2360

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rassweiler, A., Costello, C., Hilborn, R., Siegel, D.A., 2014. Integrating scientific guidance into marine spatial planning. Proc. R. Soc. B Biol. Sci. 281, 20132252.

Rooney, S.C., Rooper, C.N., Laman, E.A., Turner, K.A., Cooper, D.W., Zimmermann, M., 2018. Model-based essential fish habitat definitions for Gulf of Alaska groundfish species.

Rooper, C.N., Sigler, M.F., Goddard, P., Malecha, P., Towler, R., Williams, K., Wilborn, R., Zimmermann, M., 2016. Validation and improvement of species distribution models for structure-forming invertebrates in the eastern Bering Sea with an independent survey. Mar. Ecol. Prog. Ser. 551, 117–130. https://doi.org/10.3354/meps11703

Schirripa, M.J., Goodyear, C.P., Methot, R.M., 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. ICES J. Mar. Sci. 66, 1605–1613.

Smart, T.I., Duffy-Anderson, J.T., Horne, J.K., Farley, E.V., Wilson, C.D., Napp, J.M., 2012. Influence of environment on walleye pollock eggs, larvae, and juveniles in the southeastern Bering Sea. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 196–207. https://doi.org/10.1016/j.dsr2.2012.02.018

Stabeno, P.J., Kachel, N.B., Moore, S.E., Napp, J.M., Sigler, M., Yamaguchi, A., Zerbini, A.N., 2012. Comparison of warm and cold years on the southeastern Bering Sea shelf and some implications for the ecosystem. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 31–45. https://doi.org/10.1016/j.dsr2.2012.02.020

Teal, L., Marras, S., Peck, M., Domenici, P., 2018. Physiology-based modelling approaches to characterize fish habitat suitability: Their usefulness and limitations - ScienceDirect. Estunarine Coast. Shelf Sci. 201, 56–63.

Thorson, Ianelli, J.N., Munch, S.B., Ono, K., Spencer, P.D., 2015. Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance. Can. J. Fish. Aquat. Sci. 72, 1897–1915. https://doi.org/10.1139/cjfas-2014-0543

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

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

Thorson, J.T., 2017. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Can. J. Fish. Aquat. Sci. 75, 1369–1382. https://doi.org/10.1139/cjfas-2017-0266

Thorson, J.T., Barnett, L.A.K., 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74, 1311–1321. https://doi.org/10.1093/icesjms/fsw193

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

Wood, S.N., 2006. Generalized additive models: an introduction with R, 1st ed. Chapman and Hall/CRC Press, Boca Raton, FL.

Zhang, J., Woodgate, R., Mangiameli, S., 2012. Towards seasonal prediction of the distribution and extent of cold bottom waters on the Bering Sea shelf. Deep Sea Res. Part II Top. Stud. Oceanogr., Understanding Ecosystem Processes in the Eastern Bering Sea 65–70, 58–71. https://doi.org/10.1016/j.dsr2.2012.02.023

Table 1 – Comparison of for seventeen species analyzed and four candidate models (the model with lowest AIC is indicated in bold); Temp: a quadratic effect of local bottom temperature on numbers-density and average biomass; ColdPool: a spatially varying effect of annual cold-pool extent on numbers-density and average biomass; None: neither temperature nor cold-pool effects; Both: both temperature and cold-pool effects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | None | Temp | ColdPool | Both |
| *Gadus chalcogrammus* | 266.9 | 76.5 | 155.6 | **0.0** |
| *Gadus macrocephalus* | 572.0 | 163.1 | 392.1 | **0.0** |
| *Hippoglossoides elassodon* | 196.0 | 0.4 | 164.4 | **0.0** |
| *Chionoecetes opilio* | 28.8 | 1.9 | 25.4 | **0.0** |
| *Hippoglossus stenolepis* | 232.2 | 83.5 | 154.0 | **0.0** |
| *Limanda aspera* | 88.5 | **0.0** | 86.4 | 3.2 |
| *Pleuronectes quadrituberculatus* | 79.7 | 28.1 | 34.8 | **0.0** |
| *Chionoecetes bairdi* | 1.4 | 7.6 | **0.0** | 5.7 |
| *Podothecus accipenserinus* | 230.8 | 8.6 | 176.1 | **0.0** |
| *Atheresthes stomias* | 395.3 | 35.0 | 300.3 | **0.0** |
| *Hyas coarctatus* | 27.8 | 2.4 | 23.0 | **0.0** |
| *Myoxocephalus polyacanthocephalus* | 83.5 | 11.7 | 55.3 | **0.0** |
| *Lycodes palearis* | 129.2 | 1.1 | 98.3 | **0.0** |
| *Myoxocephalus jaok* | 103.9 | 23.6 | 74.2 | **0.0** |
| *Hyas lyratus* | 0.1 | **0.0** | 2.9 | 3.4 |
| *Paralithodes camtschaticus* | 45.8 | 7.2 | 35.2 | **0.0** |
| *Lycodes brevipes* | 4.1 | **0.0** | 7.5 | 3.0 |

Figure 1: Maps showing local bottom temperatures in each year 1982-2017 for each of 200 “knots” used within VAST to approximate spatio-temporal variation in density for each modeled species (see colorbar left in bottom-right panel; blue represents low temperatures and red represents high temperatures).

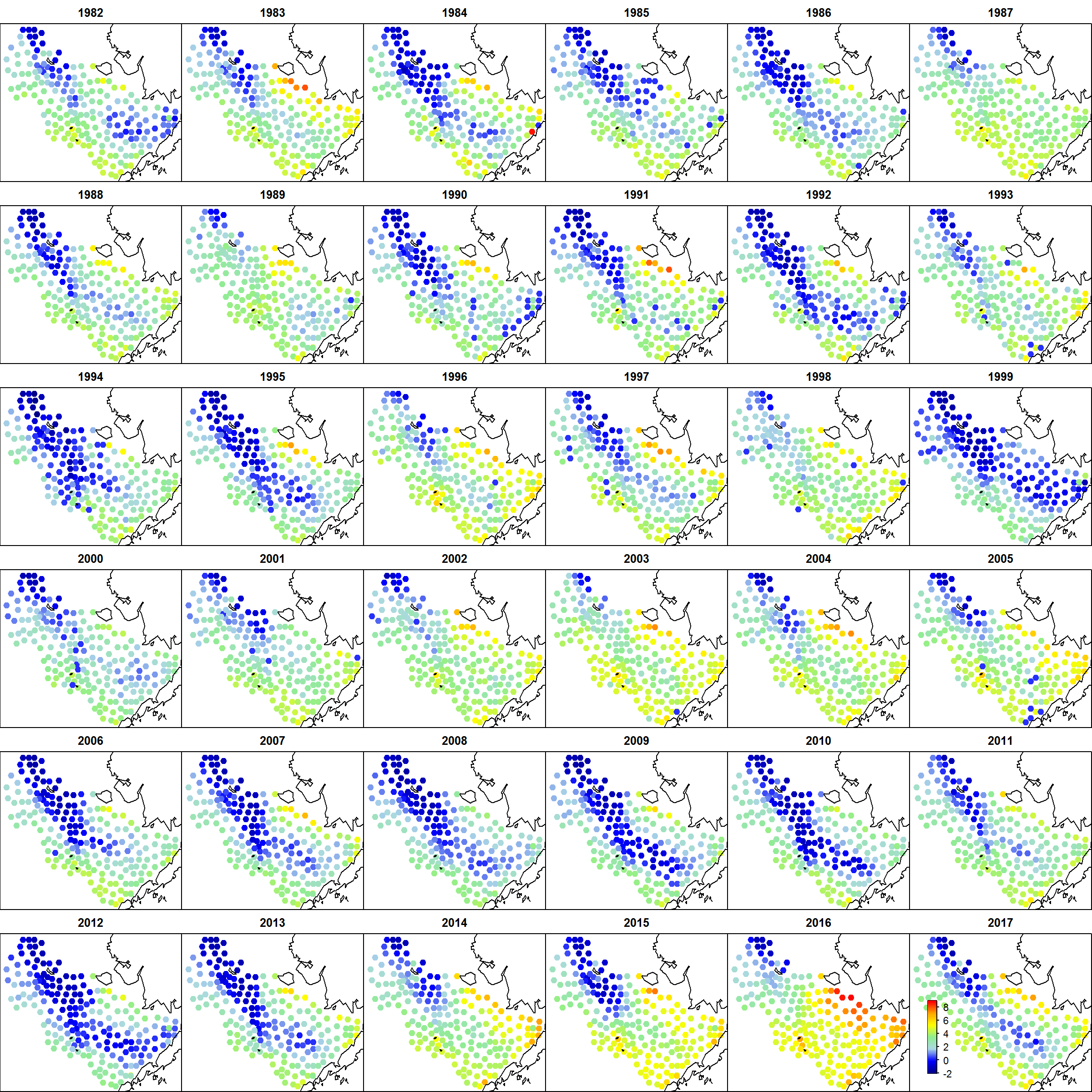


Figure 2: Time-series showing extent of cold-pool in the Eastern Bering Sea for modeled years, 1982-2017 (x-axis), where the cold-pool is measured as the area with bottom temperatures at or below . The time-series is labeled both for the measured value (left-hand y-axis label) and the standardized value (right-hand y-axis label), where the standardized value is included in the model.

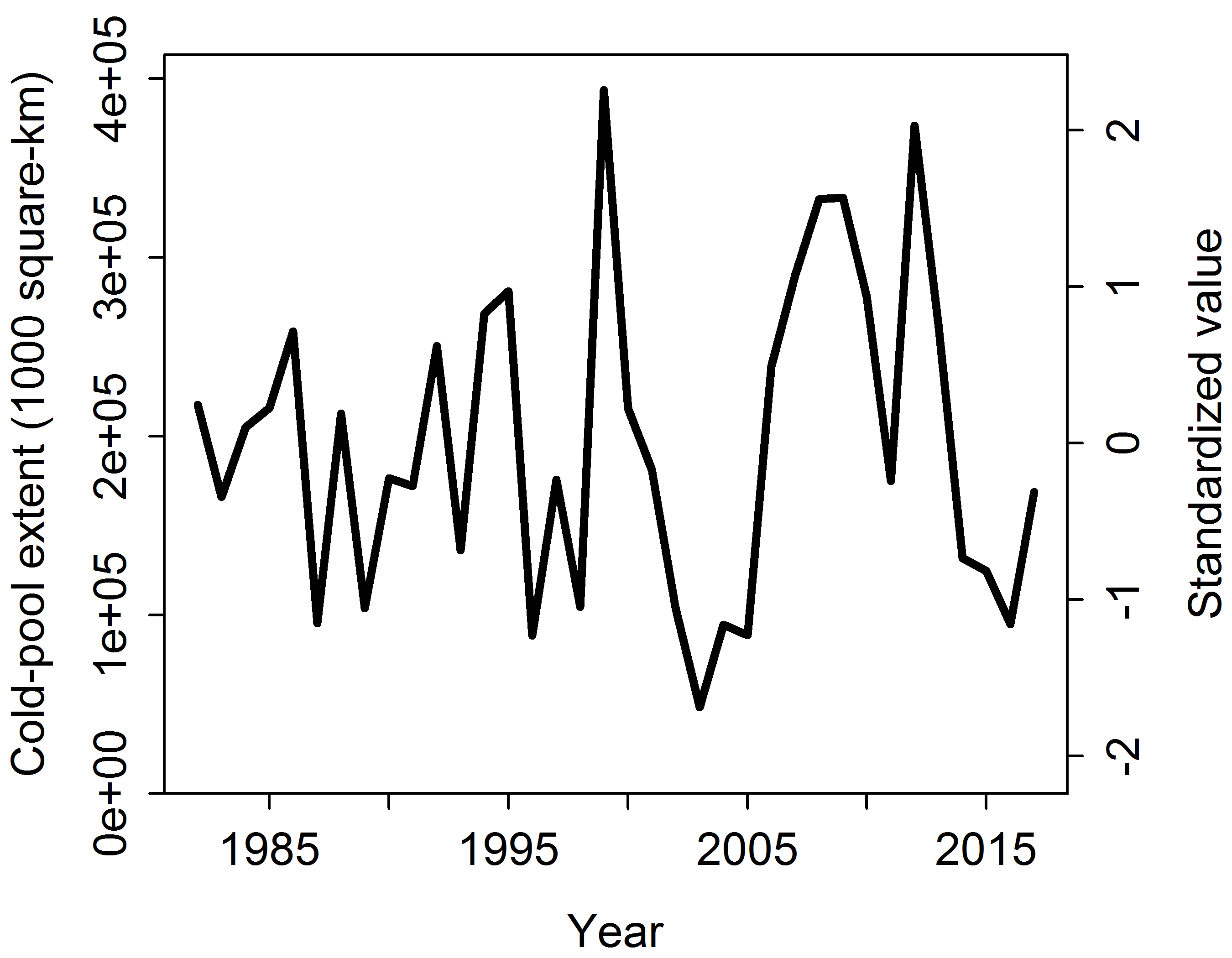


Figure 3 – Illustration of the spatially varying coefficient linking annual cold-pool extent to either numbers density (left column) or average-weight (right column) in the Poisson-link delta model used to predict biomass for each species in the Eastern Bering Sea. Results are shown for the model including a cold-pool effect but not bottom temperature. Cold-pool extent is standardized to have a mean of zero and standard deviation of one prior to use and all results are using a Poisson-link delta model such that cold-pool extent has a linear effect on log-biomass density; consequently, a location with a coefficient of indicates an approximately 10% increase in expected density for every 1 standard deviation (87,504 ) increase in cold-pool extent. Note that the color-scale differs for each species (rows, labeled on left) and linear predictor (see color-bar for scale in each panel).

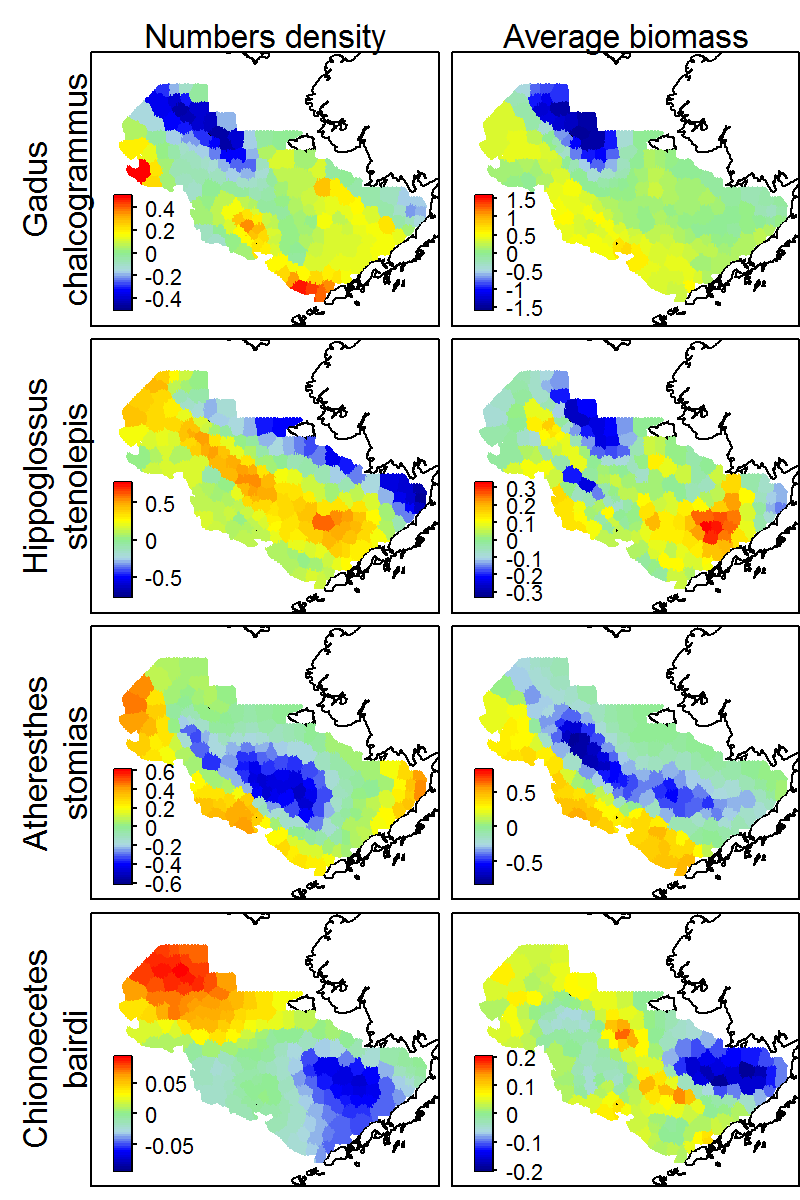


Figure 4: Summary of the standard deviation of variation in log-density caused by a spatially varying link to cold-pool extent (1st row), a quadratic response to bottom temperature (2nd row), or residual spatio-temporal variation (3th row) as impacting either numbers density (1st column) or average biomass (2nd column). Each panel shows a boxplot (line: mean; box: interquartile range; whiskers: the furthest point within 1.5 times the interquartile range from the mean) summarizing the estimated standard deviation across all seventeen analyzed species for a given model (None: neither temperature nor cold-pool; Both: both temperature and cold-pool responses). All panels also list the average standard deviation across all seventeen species above each boxplot, and the bottom row also lists the average reduction in spatio-temporal variance relative to the model without covariates (where, e.g., 50% would indicate that covariates explain half of spatio-temporal variance). Note the different y-axis scale for each row.

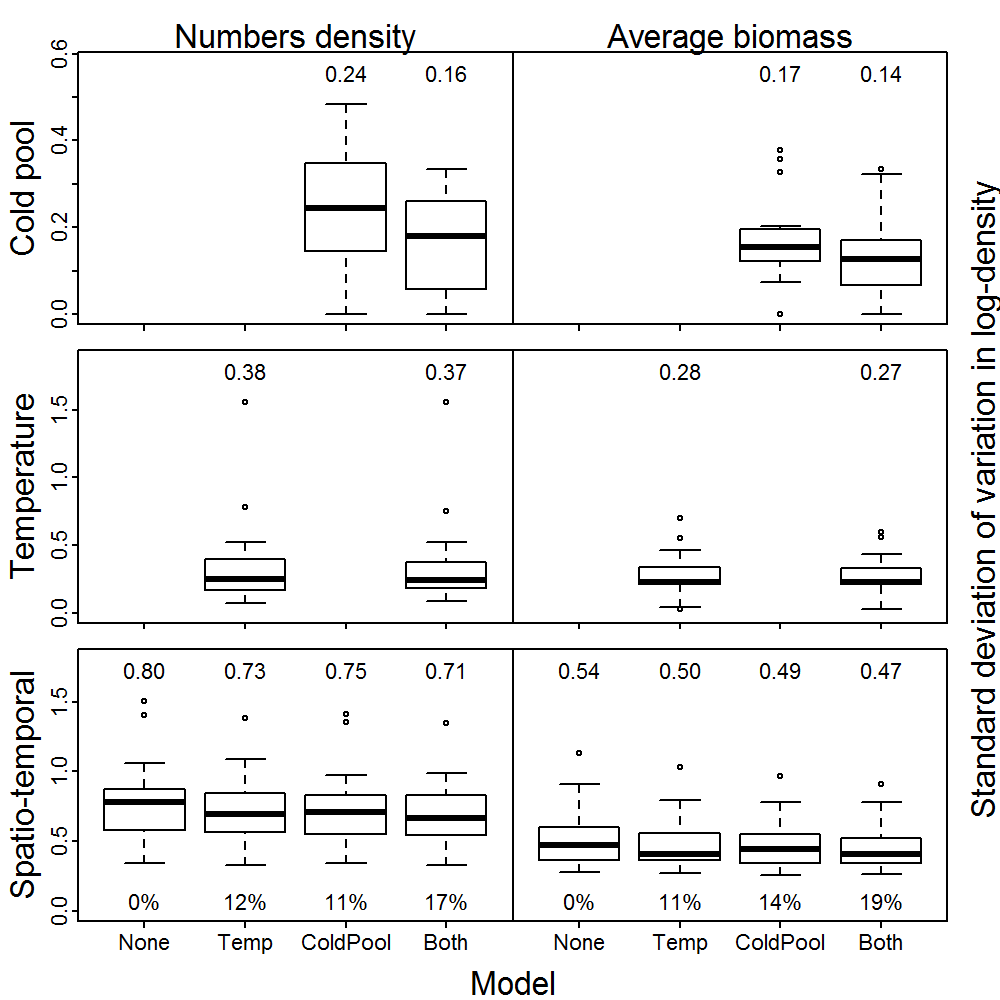


Figure 5: Estimates of northward center-of-gravity (y-axis; note different scale for each species) for each year 2012-2017 (x-axis) for each of twenty species (panels) and four models (line: estimate; shaded area: +/- one standard error; see color labels in top-left panel) when fitting to data 1982-2014 and forecasting distribution in 2015-2017. The dashed lines show estimates when fitting to all data (2012-2017) for each model, which are generally very similar. Each panel includes the average error (computed as the difference between forecasts and estimates using all data when averaging across models) for each of four models as well as a “persistence” forecast (i.e., forecasting that the distribution does not change after 2014).

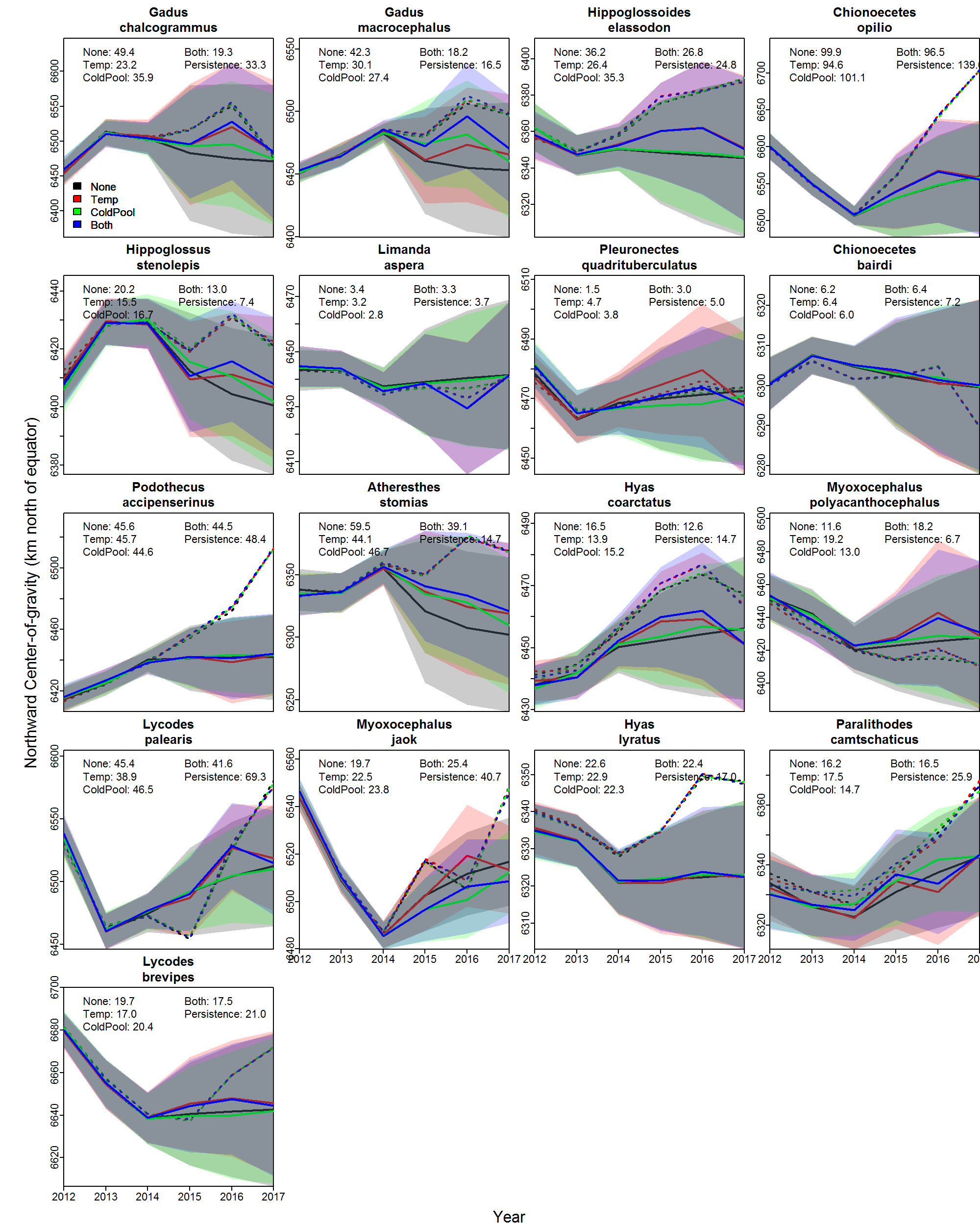


Fig. 6 – Distribution of error for forecasts of northward center of gravity (computed as the difference between forecasts and estimates using all data when averaging across models) for all twenty species when fitting to data 1982-2014 and forecasting distribution in 2015 (left panel), 2016 (middle panel) or 2017 (right panel) using four models or a “persistence” forecast (see Fig. 5 caption for details). The top of each panel lists the median absolute error (a value close to zero is better) and bias (a value close to zero is better) for each forecast model.

