ICES Journal of Marine Science

ICES Journal of Marine Science (2022), 79(1), 117–128. https://doi.org/10.1093/icesjms/fsab247

## Original Article

# Development and simulation testing for a new approach to density dependence in species distribution models

James T. Thorson *

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

**Corresponding author: tel: 206-526-4517; e-mail: James.Thorson@noaa.gov*

Thorson, J. T. Development and simulation testing for a new approach to density dependence in species distribution models. – ICES Journal of Marine Science, 79: 117–128.

Received 22 October 2021; revised 4 November 2021; accepted 8 November 2021; advance access publication 15 December 2021.

Density dependence is included in many population–dynamics models, but few options exist within species distribution models (SDMs). One option for density-dependence in SDMs proceeds by including an independent time-series of population abundance as covariate using a spatially varying coefficient (SVC). We extend this *via* three alternative approaches that replace the independent time-series with information available within the SDM. We recommend the "intermediate complexity" approach that estimates a SVC responding to median abundance in each time; this SVC indicates whether a given location has a smaller- or greater-than-average sensitivity to changes in median abundance. We next develop a reaction–advection–diffusive simulation model, wherein individuals avoid habitats that exceed a threshold in local density. This movement model results in an estimated SVC that is negatively correlated with the average spatial distribution. Finally, we show that a SVC can be identified using bottom trawl data for four species in the eastern Bering Sea from 1982 to 2019. We conclude that the common "basin-model" for animal movement results in an ecological teleconnection, wherein population depletion or recovery at one locations will affect resulting dynamics at geographically distant habitats, and that this form of density dependence can be detected using SDMs.

**Keywords:** basin model, density-dependent habitat selection, habitat preference, reaction–advection–diffusion, species distribution model, Vector Autoregressive Spatio-Temporal (VAST).

## Introduction

Climate changes is causing rapid and easily observed shifts in spatial distribution for many taxa worldwide, often towards poles and/or higher elevations (Parmesan, 2006; Stevenson and Lauth, 2019). This has prompted increased interest in using spatial distribution as a sensitive indicator of population vulnerability (Keith *et al.*, 2013) and climate impacts (Pinsky *et al.*, 2013). However, many top-down and bottom-up processes impact spatial distribution simultaneously (Planque *et al.*, 2011). As a result, ecologists need tools to identify mechanisms causing shifts in spatial distribution.

In particular, local competition will often cause habitat quality or preference to decrease as population densities increase, both due to decreasing prey resources, saturated refuge from predators, or other forms of behavioural interference (see recent discussion in Avgar *et al.*, 2020). As a result, an increase in total abundance is sometimes associated with individuals colonizing new habitats and a resulting increase in total area occupied; this is called the "basin model for biogeography" (MacCall, 1990). Alternatively, an increase in abundance may instead be associated with an increase in density within existing habitat, termed a "proportional density model" (Bartolino *et al.*, 2011). An analysis of survey data for 120 groundfish stocks generally supported the "proportional density model," wherein the relationship between abundance and area occupied was highly variable and only weakly positive on average (Thorson *et al.*, 2016). However, the same analysis showed that high-latitude gadids had a stronger relationship on average than other taxa, and a strong relationship between population range and abundance has also been documented in several high-profile commercial species (MacCall, 1990; Walters and Maguire, 1996).

To better study shifts in spatial distribution, ecologists have developed a wide range of statistical and mechanistic models to simulate and estimate species densities across a landscape. Methods include "correlative" species distribution models (SDMs) that are

typically fitted as a statistical regression (e.g. Warton and Shepherd, 2010) and reaction–advection–diffusion (RAD) models that explicitly approximate habitat-specific productivity and preferences (Hanks *et al.*, 2015). Correlative SDMs are now widely used for stock, habitat, climate, and ecosystem research in fishes and other taxa (Elith *et al.*, 2010), while RAD models are less widely used, presumably due to a lack of easy-to-use software to fit them to data (although see Thorson *et al.*, 2021a).

Despite this interest in distributional shifts and the broad use of SDMs in applied ecology and fisheries management, there are surprisingly few applied tools to include density-dependence within correlative SDMs. As notable exceptions, previous studies have combined field samples of local density with an independent time-series of total abundance from stock assessments to estimate the spatially varying response of local density to changes in stock abundance (Bacheler *et al.*, 2009, 2012; Bartolino *et al.*, 2011; Ciannelli *et al.*, 2012). This approach can be implemented using widely used generalized additive models (GAMs), and has illustrated the combined effect of density-dependent and environmental drivers on the spatial distribution of groundfishes. However, it also requires treating stock-assessment estimates of total abundance as a covariate that is known without error. This in turn causes SDM results to depend upon subjective and opaque choices made during the stock assessment, which has been criticized in other contexts (Dickey-Collas *et al.*, 2014; Brooks and Deroba, 2015). Additionally, the performance of this GAM approach has not been tested previously using simulated data.

For these various reasons, we seek to develop methods to include density-dependent mechanisms within correlative SDMs without requiring an independent time-series of population abundance (from stock assessments or otherwise). To do so, we first introduce three alternative statistical approaches fitted only to samples of local population density, which vary in terms of conceptual and computational complexity. We then describe in detail our implementation of the "intermediate complexity" approach, which includes the temporal main effect as a covariate with a spatially varying response. Next, we introduce a RAD simulation model, and use this to conduct the first (to our knowledge) simulation test of the performance of a density-dependent SDM. We specifically explore whether (1) the estimation model can identify the spatial effect of density dependence, while (2) also appropriately discriminating between density-dependent and –independent simulation scenarios. Finally, we apply our estimation model to data for four species in the eastern Bering Sea, which show a variety of density-dependent responses. All three statistical approaches are implemented in the publicly available R package VAST (Thorson and Barnett, 2017; Thorson, 2019b) starting with release number 3.8.0 (https://github.com/James-Thorson-NOAA/VAST/) to facilitate future exploration and application.

## Methods

### Summarizing the spatio-temporal SDM

We quickly summarize notation for a conventional SDM, before adapting this notation to introduce three alternative implementations for density dependence. We specifically describe a generalized linear mixed model (GLMM) that fits to samples of population biomass $b_i$ for each sample $i$ occurring at location $s_i \in \mathcal{D}$ within a spatial domain $\mathcal{D}$ and at discrete time intervals $t_i \in \{t_1, t_2, \ldots, n_t\}$. In the following we envision that samples measure biomass and hence use a Tweedie distribution for residual variation:

$$b_i \sim Tweedie\left(a_i d_i, \phi, p\right), \tag{1}$$

where $a_i$ is the area-swept by each sample and $d_i$ is the expected biomass density such that $\mathbb{E}(b_i) = a_i\, d_i$, and $\phi$ and $p$ controls the magnitude and heteroscedasticity of variance, $\mathbb{V}(b_i) = \phi d_i^p$ (Foster and Bravington, 2013). The following could easily be adapted to instead use a distribution that is customized for presence–absence or count data.

We then include a main effect of time ($\beta$) and space ($\omega$), an interaction between space and time ($\varepsilon$), as well as a spatially varying response to covariates affecting density:

$$\log(d_i) = \underbrace{\beta(t_i)}_{time} + \underbrace{\omega(s_i)}_{space} + \underbrace{\varepsilon(s_i, t_i)}_{interaction} + \underbrace{\sum_{k=1}^{n_k} \left(\gamma_k + \chi_k(s_i)\right) x_k(s_i, t_i)}_{covariates} \tag{2}$$

where covariates $x_k(s_i, t_i)$ measure the value of $n_k$ processes at the location $s_i$ and time $t_i$ associated with sample $i$, and $\gamma_k + \chi_k(s)$ represents the spatially varying response to that covariate. In the following, we treat vectors $\omega$, $\varepsilon(t)$, and $\chi_k$ as zero-mean Gaussian Markov random fields (GMRF) using a stochastic partial differential equation (SPDE) approximation to calculate their probability given a sparse precision matrix that approximates a Matérn correlation function (Lindgren et al., 2011). Treating $\chi_k$ as a spatial random effect is a special case of a "random slope" regression or varying coefficient model (Hastie and Tibshirani, 1993), and its use in SDMs was explored by Thorson (2019a). This spatially varying coefficient (SVC) approach has been adapted for other purposes (Barnett *et al.*, 2021), and builds upon previous studies approximating a spatially varying response as a spline within a GAM (Bacheler *et al.*, 2009, 2012; Bartolino *et al.*, 2011; Ciannelli *et al.*, 2012).

In particular, changes in population density $d(s, t)$ at location $s$ prior to time $t$ could affect density $d(s^*, t)$ at any other location due to movement of resources, competitors, and predators among locations. Estimating the effect of density at each of $S$ locations on productivity at every other location involves $S^2$ potential associations, and estimating these becomes infeasible given many locations. Therefore, SDMs are often specified such that covariates $x(s, t)$ only affect local density at that single location. In some cases, however, $S^2$ nonlocal associations can be approximated in two steps: by (1) reducing dimensionality from a density matrix $\mathbf{D}$ (with dimension $S$ by $T$, where $T$ is the number of times) to one or a few regional indices $\mathbf{v}(f)$ with a value $v(f, t)$ in each year, and then (2) identifying how change in density at each location is associated with each index $\mathbf{v}(f)$. Step #1 is often implemented using "rank reduction" methods such as empirical orthogonal functions (EOF; Thorson *et al.*, 2020), and Step #2 using a SVC model (Thorson, 2019a). These two steps collectively describe an "ecological teleconnections," wherein dynamics at multiple locations are correlated across space due to their dependence upon the few regional indices (Litzow *et al.*, 2018). In species with strong density dependence, total abundance might be a parsimonious choice for $\mathbf{v}(f)$, while the spatially varying response represents the local effect of changes in total abundance (in Step #2). For this reason, we interpret a spatially varying response to population abundance as one plausible form of "ecological teleconnection." Modelling a local impact of total abundance is a convenient way to include density-dependence within an SDM, but clearly has different

rationale and structure than a mechanistic theory of animals selection among habitats.

All SDMs explored here are fitted using the Vector Autoregressive Spatio-Temporal (VAST) package using release number 3.8.0 in the R statistical environment version 4.0.2 (R Core Team, 2020). In this study, we assume that $\varepsilon(t)$ is independent for any pair of years and that temporal main effects $\beta$ are freely estimated as fixed effects. This model structure corresponds to the default "index standardization" settings in VAST, as discussed in detail elsewhere (Thorson, 2019b), although users could instead specify an autoregressive structure for $\varepsilon(t)$ and/or $\beta$ to allow future dynamics to be forecasted. We also estimate as fixed effects the Tweedie variance parameters $\phi$ and $p$, the variance and decorrelation-rate parameters for the GMRFs, and any covariate response parameters $\gamma_k$. We use maximum likelihood (ML) to identify the value of fixed effects while integrating across the probability of random effects (Kristensen *et al.*, 2016), and then maximize the joint likelihood with respect to random effects (conditional upon these ML estimates) to identify Empirical Bayes predictions for random effects. We then use these ML estimates and empirical Bayes predictions to predict biomass–density at a set of "extrapolation-grid" cells $g$ in each modelled year $t$ and calculate total biomass as the sum of these predictions, while using the epsilon bias-correction method to correct for retransformation bias (Thorson and Kristensen, 2016).

## Three ways to include density dependence

We next summarize three alternative approaches to including density dependence without using an independent time-series of abundance (e.g. from a stock assessment), which we introduce in order from simple to complex. In all three approaches, we use a zero-centered response to a time-series representing total or median abundance (i.e. $\gamma_k = 0$ for the associated response) because estimating a non-zero average response would be confounded with the temporal main effect $\beta(t)$. The model formulation [Equation (2)] is invariant to a scale transformation of density and abundance (i.e. changing units from kilograms to tons), but is not invariant to changes in the "centering" of a given index (i.e. subtracting a covariate that uses a SVC response by a constant will affect resulting parameter estimates). Future developments could eliminate this sensitivity by estimating a correlation between spatial component $\omega$ and SVC response $\chi_k$, although we do not explore this topic further here.

## Two-stage

As the simplest approach, an analyst could fit this SDM without including density dependence, extract the estimate of total biomass $I_t$ in each year, and then include centered log-biomass as a covariate $x_1(s,\ t) = \log(I_t) - n_t^{-1} \sum_{t=1}^{n_t} \log(I_t)$ when re-fitting the model. In this case, the spatially varying response $\chi_1(s)$ represents the log–log response of biomass-density at location $s$ to changes in total biomass, e.g. where $\chi_1(s) = 0.1$ means that a density increases approximately 10% faster than the median location $s$ when total biomass doubles. We include centered log-biomass (i.e, . subtracting $\log(I_t)$ by its mean $n_t^{-1} \sum_{t=1}^{n_t} \log(I_t)$ prior to using it as covariate) such that the spatial term $\omega(s)$ represents the distribution in an "average year" when $x_1(s,\ t) = 0$. This approach is similar to past efforts modelling density dependence using a GAM, but replacing the stock-assessment estimate of abundance with estimated abundance from an initial fit to available sampling data.

## Temporal effect as covariate

As an approach of intermediate complexity, we note that $\beta(t) + \bar{\varepsilon}(t)$, where $\bar{\varepsilon}(t) = n_s^{-1} \sum_{s'=1}^{n_s} \varepsilon(s',\ t)$, controls the median density across space. Changes in $\beta(t) + \bar{\varepsilon}(t)$, therefore, represents inter-annual changes in total abundance after controlling for the effect of other covariates. We propose fitting the model while specifying that a covariate is equal to a centered index of median density, $x_1(s,\ t) = (\beta(t) + \bar{\varepsilon}(t)) - n_t^{-1} \sum_{t=1}^{n_t} (\beta(t) + \bar{\varepsilon}(t))$, where we again use a centered index such that $\omega(s)$ represents distribution in an average year. The spatially varying effect of this covariate, $\chi_k(s)$, again represents each location's response to variation in population size from year-to-year. Future research could additionally add the effect of other covariates to this index of median density prior to using it to model density dependence, and this would propagate the effect of covariates that might increase abundance and indirectly cause a density-dependent response. We do not explore this topic further here, but recommend it as a potentially useful extension.

## Lagrange multiplier

Finally, as the most complex approach, we introduce a new auxiliary parameter $\iota(t)$ for each year. We then specify that the covariate is equal to this value, $x_1(s,\ t) = \iota(t) - n_t^{-1} \sum_{t=1}^{n_t} \iota(t)$ while also introducing a constraint that it must be approximately equal to total biomass:

$$\log(I_t) \sim Normal\left(\iota(t), \sigma_\iota^2\right), \tag{3}$$

where $\sigma_\iota^2$ is the magnitude of a penalty that is specified a priori (rather than estimated), and this penalty ensures that $\iota(t)$ approximately equals $\log(I_t)$. Constructing the model in this way ensures that the estimated (i.e. input) value for covariate $\iota(t)$ is approximately equal to the derived (i.e. output) value for log-abundance $\log(I_t)$, while preserving the calculation of the joint likelihood as a directed acyclic graph. For given values of all random and fixed effects except $\beta(t)$, this penalty could be interpreted as defining a hyperdistribution for intercepts $\beta(t)$. Alternatively, in the limit as $\sigma_\iota^2 \to 0$, this penalty could be interpreted as a Lagrange multiplier to ensure that $\beta(t)$ approaches the value such that $\iota(t) \to \log(I_t)$.

## Contrasting these three approaches

The two-stage approach does not require any specialized software and is simple to explain and interpret. However, it treats $I_t$ as if it were known without error, and therefore, does not propagate estimation uncertainty. By contrast, both "intermediate complexity" and Lagrange-multiplier approaches propagate uncertainty about the estimated value of this covariate. In the "intermediate complexity" approach, $\beta(t) + \bar{\varepsilon}(t)$ will not be perfectly correlated with total biomass, and hence the estimate of spatially varying responses is not as easy to interpret as the two-stage model. By contrast, for the "Lagrange multiplier" approach, $\iota(t)$ is equivalent to $\log(I_t)$ as the penalty becomes large, and this ensures that the spatially varying response can be interpreted as a response per proportional change in total abundance.

Despite the conceptual advantages for the "Lagrange multiplier" approach, we have also found several disadvantages, i.e. that:

1. results from this approach are sensitive for some species to the exact value used for $\sigma_t^2$;
2. high values of $\sigma_t^2$ will result in a failure of the Laplace approximation; and
3. fits will sometimes shrink estimates of density in otherwise unsampled areas towards values that result in $\log(I_t)$ being a good predictor of spatial dynamics, but which otherwise seem implausible.

Based on these concerns (summarized in Table 1), we explore the "intermediate complexity" approach within a simulation design, and also compare results from both "two-stage" and "intermediate complexity" approaches for four species within a case-study application.

## Reaction-Advection-Diffusion (RAD) simulation model

We first test the density-dependence SDM *via* application to data generated from a RAD simulation model (see Appendix A for computational details). We specifically explore performance for two scenarios that either include or exclude density-dependent habitat selection. To our knowledge, this is the first simulation experiment testing the performance of a spatially varying response to abundance as approximation to density-dependent habitat selection.

The RAD simulation model involves simulating numerical abundance $n(s,\ u,\ y)$ at each location $s$, month $u$, and year $y$ that follows an exponential growth rate $r_y$ during each of 12 months over 70 years (840 intervals total). The first 25 years are treated as a "burn-in" period (to ensure that spatial dynamics in the first year after this burn-in includes reasonable variation around the initial distribution) and the remaining 45 years result in sampling data (e.g. representing data collected from 1975 to 2019). We restrict the simulation to the eastern and northern Bering Sea (see Appendix A, Figure A1 to identify locations discussed in the text) and discretize this spatial domain into 100 square cells each with dimension 100 km × 100 km. We simulate bathymetry based upon reconstructed smoothsheets (Zimmermann and Benson, 2013; Zimmermann and Prescott, 2018), and bottom temperatures based on the Bering10K Regional Ocean Modelling System (Kearney *et al.*, 2020).

For each of 50 simulation replicates, we simulate a growing population by specifying an annual growth rate $r_y$ (with units $yr^{-1}$) that is typically positive, but which includes autocorrelated variation in growth rate to generate contrast among simulation replicates. We then simulate individual movement based on a habitat-preference function that depends upon local bathymetry, bottom temperatures, as well as a response to local densities at the beginning of each month. We contrast results from this "density-dependent movement" scenario with a scenario that eliminates density dependence from the habitat preference function. Using a habitat-preference function to approximate individual movement builds upon recent research (Thorson *et al.*, 2021a, b). However, this is the first study (to our knowledge) that includes a response to local densities in the habitat-preference function to approximate density-dependent habitat selection.

We initialize abundance $n(s,\ u,\ y)$ in the first month $u$ and year $y$ from the stationary distribution for habitat utilization (as calculated hypothetically given zero densities), and then distribute 10 million individuals over the spatial domain according to this stationary habitat utilization. We project dynamics forward for each month of the 25 burn-in years and again for the 45 years of sampling data, and record effective area occupied in all years (Thorson *et al.*, 2016). Simulated sampling follows a Tweedie distribution assuming that each individual has average biomass of 1.5 kg and a coefficient of variation in weight of 1.0 and where each tow samples 0.01 sq km. We simulate 100 samples occurring in July of each year, yielding a total sample size of 4500 density samples for each replicate. Given our simulation design, we expect density-dependent habitat selection to result in a positive correlation between effective area occupied and total abundance.

For each simulation replicate, we fit simulated samples using the spatio-temporal SDM using the "intermediate complexity" approach for density dependence. We fit this model using 100 spatial knots that are distributed in proportion to 2000 extrapolation-grid locations, which are in turn distributed evenly throughout the eastern and northern Bering Sea survey domains. We include density dependence as the only covariate, such that estimates of spatial and spatio-temporal model components approximate the combined effect of bathymetry and temperature. We record the estimated spatial variable $\omega(s)$ as well as the spatially varying response $\chi_1(s)$, and repeat this for all replicates of both density-dependent and -independent scenarios.

A well-performing estimator will estimate a substantial response to changes in total abundance in the scenario that includes density dependence, and conversely will estimate a small or nonexistent response to total abundance in the density-independent scenario. We, therefore, contrast the estimated magnitude of the density-dependent response between these two scenarios. The simulation model starts with the majority of the population in deeper habitat, but it subsequently colonizes shallowed habitat as abundance increases. By contrast, the estimation model estimates a spatial component $\omega$ (representing the average spatial distribution across years) and the density-dependent response $\chi_1$ (representing the areas with a faster- or slower-than-average increase in density as abundance increases). The estimated spatial component will not exactly equal the simulated depth covariate or the initial spatial distribution, but a well-performing estimator in the density-dependent scenario will still capture the progressive expansion into new habitat. This will then result in a negative correlation between the estimated spatial component $\omega$ and the density-dependent response $\chi_1$. We, therefore, record this correlation in both density-dependent and -independent scenarios. Finally, we extract the average across years of the density-dependent preference [the third term on the right-hand-side of Equation (A2)], and calculate the correlation between this and the density-dependent response $\chi_1$. A well-performing estimation model will have a high correlation, indicating that the density-dependent response is a good proxy for the average effect of density-dependent preference.

## Case study demonstration

We next fit both the "two-stage" and "intermediate complexity" approaches to survey data for four species with contrasting life-history. We specifically use samples from a systematic bottom trawl survey conducted in the eastern Bering Sea from 1982 to 2019 by the Alaska Fisheries Science Center, using a large mesh (83–112) gear and sampling approximately a hectare (0.01 sq km) calculated as the product of tow length measured from bottom sensors and net width mensuration (Lauth and Conner, 2016). We again fit this using a Tweedie distribution and 100 knots, which are distributed in

**Table 1.** A quick summary of the benefits and drawbacks of the three alternative approaches to including density dependence within a correlative SDM.

| Name of approach | Benefits | Drawbacks |
|---|---|---|
| Two-stage | • Simple to explain<br>• Does not require specialized software | • Does not propagate variance |
| Intermediate complexity | • Propagates variance<br>• Numerically stable | • Index of density dependence represents median (not total) abundance |
| Lagrange multiplier | • Conceptually appealing | • Heavily shrinks density in hotspots to achieve good fit<br>• Depends upon a priori specification of penalty term |

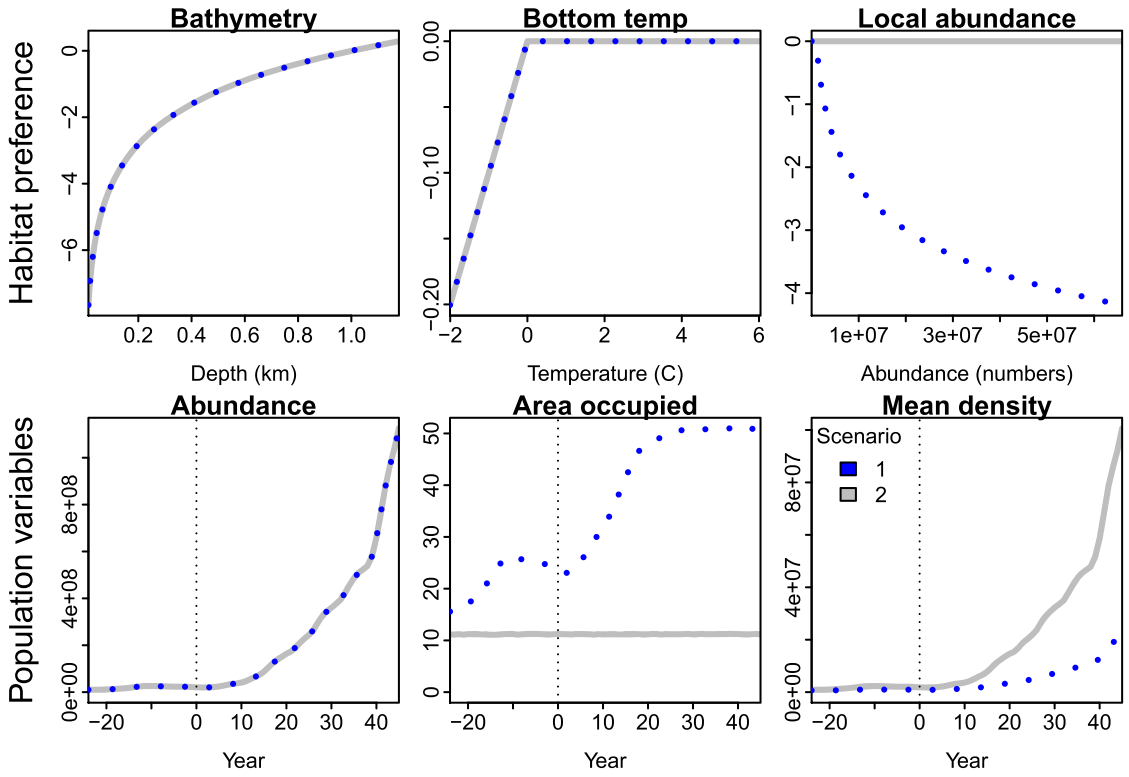


**Figure 1.** Summary of the operating model showing habitat-preference (top row) for depth, bottom temperature, and local abundance for two scenarios that include (Scenario 1: dotted blue) or do not include (Scenario 2: solid grey) a density-dependent response for habitat preference, and also showing the simulated abundance for the first replicate, the effective area occupied, and the per-capita average density that results from those two scenarios (bottom row).

proportion to the 2000 extrapolation-grid locations that are evenly distributed across the eastern Bering Sea survey domain. We include density dependence as the single covariate and compare resulting estimates of the spatially varying response between the "two-stage" and "intermediate complexity" approaches.

Finally, we explore whether the density-dependent model can generate changes in habitat utilization for each case-study species using the "intermediate complexity" approach. To do so, we project density in the absence of any residual spatio-temporal variation across the range of observed values for the temporal effect, and visualize whether habitat utilization expands or contracts across this range. Specifically, we calculate $d(s,\ t)$ in Equation (2) using estimates of $\omega(s)$ and $\chi(s)$ but fixing $\varepsilon\ (s,\ t) =\ 0$ and using values $\beta(t)$ at the minimum, maximum, or the average of these two values. This projected density isolates the effect of density dependence in explaining changes in habitat utilization given high or low abundance. We note that detecting a density-dependent response (i.e. $\chi(s) \neq 0$) is not evidence of density-dependent habitat selection *per se*, because other processes may also result in a density-dependent response (Shepherd and Litvak, 2004), but we hope that evidence of a density-dependent response will still motivate further research to understand the resulting estimates of $\chi(s)$.

## Results

The simulation model results in identical abundance in the first replicate for the scenarios with or without density dependence (Figure 1, bottom-left panel), but the density-dependent scenario results in higher area occupied (blue line in Figure 1, middle bottom panel) because local density (Figure 1, bottom-right panel) ex-

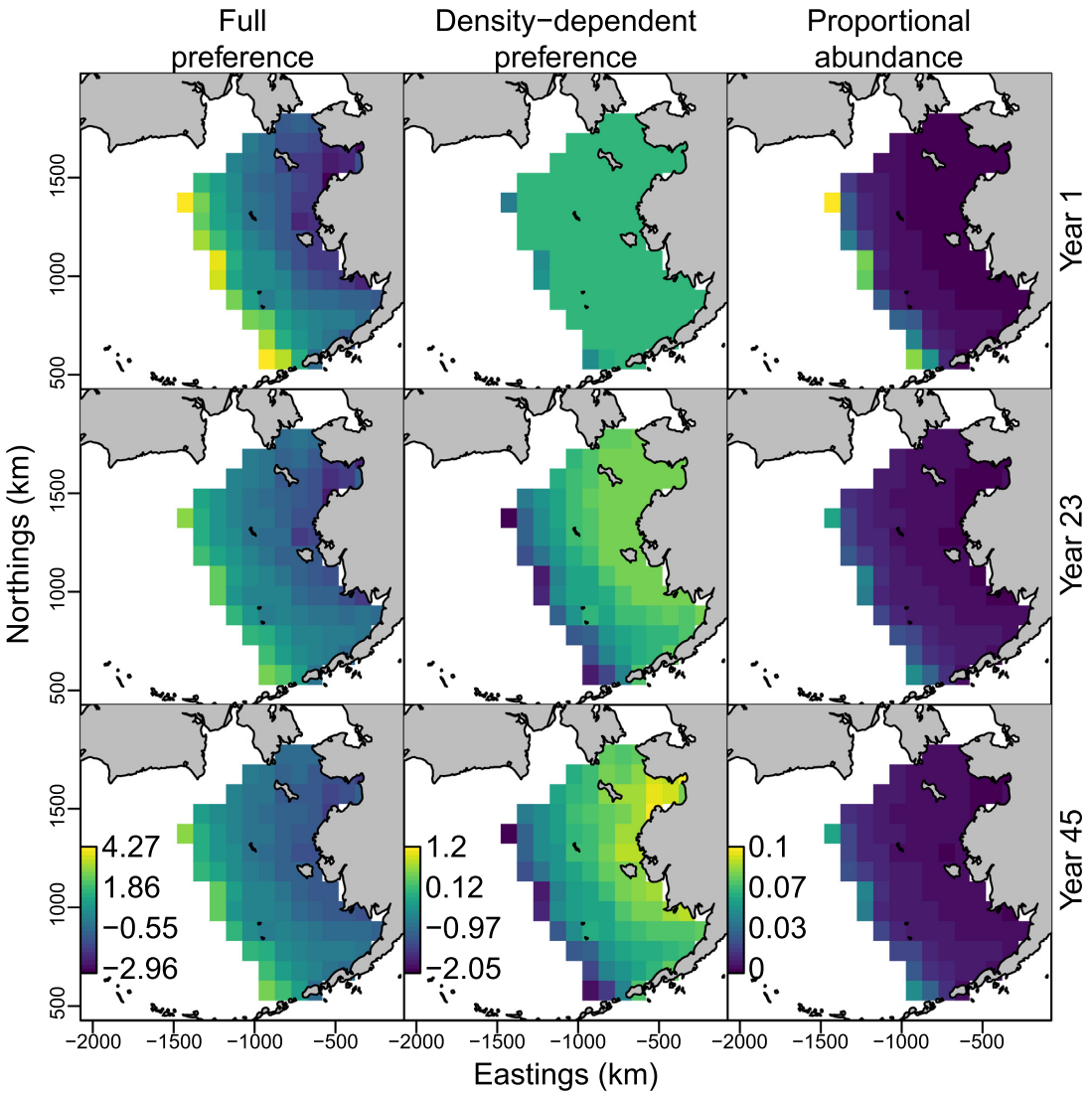


**Figure 2.** Depiction of habitat preference used to drive taxis [left column, $h(s,\ t)$ from Equation (A2)], the component of preference due to density dependence [middle column, $f_{\_}[\log(\frac{x_3(s,\ t)}{10^6})]$ from Equation (A2)], and resulting abundance expressed as a proportion of total abundance (right column) for the first (top row), middle (second row), and last (bottom row) year of the sampling data from the first replicate of the density-dependent scenario.

ceeds the threshold at which habitat preferences decrease. Specifically the effective-area occupied increases from approximately 15–50% of total area during the modelled period in the first replicate of the density-dependent scenario, but stays at approximately 10% in the density-independent scenario (with small deviations resulting from interannual variation in the bottom-temperature landscape). Inspecting maps of habitat preference for this replicate (Figure 2, left column) shows that preference is nearly seven units higher for the shelf-break than nearshore habitats in the first year of sampling (Figure 2, top row), but this difference in preference is reduced by nearly half in later years (Figure 2, middle and bottom rows). This difference results from the density-dependent component of preference [third term on right-hand-side of Equation (A2)], which is 3.25 units higher in nearshore than offshore habitats in these later years (Figure 2, middle column). As a result, nearly 10% of the total abundance in the first year is in a single cell in the northern shelf break with maximum density, whereas later years of have 3–5% of total abundance in any single cell (Figure 2, right column).

Fitting the "intermediate complexity" SDM to this same replicate of the simulation experiment estimates a spatial term, $\omega(s)$, that is nearly 3 units higher in the outer domain than Norton Sound (Figure 3, second column, and see Figure A1 for location labels), and estimates for the second replicate (Figure 3, middle row) as well as the median across replicates (Figure 3, bottom row) are qualitatively similar (although showing some differences in the exact spatial pattern). In the density-dependent simulation experiment, the estimated density-dependent response $\chi(s)$ is high northeast of St. Matthew Island, and in other replicates is similarly high in Bristol Bay or other portions of the middle domain (Figure 3, third column). The average over time of the density-dependent component of preference (Figure 3, first column) in the RAD simulation model is lower in the outer domain than middle and inner domain, which

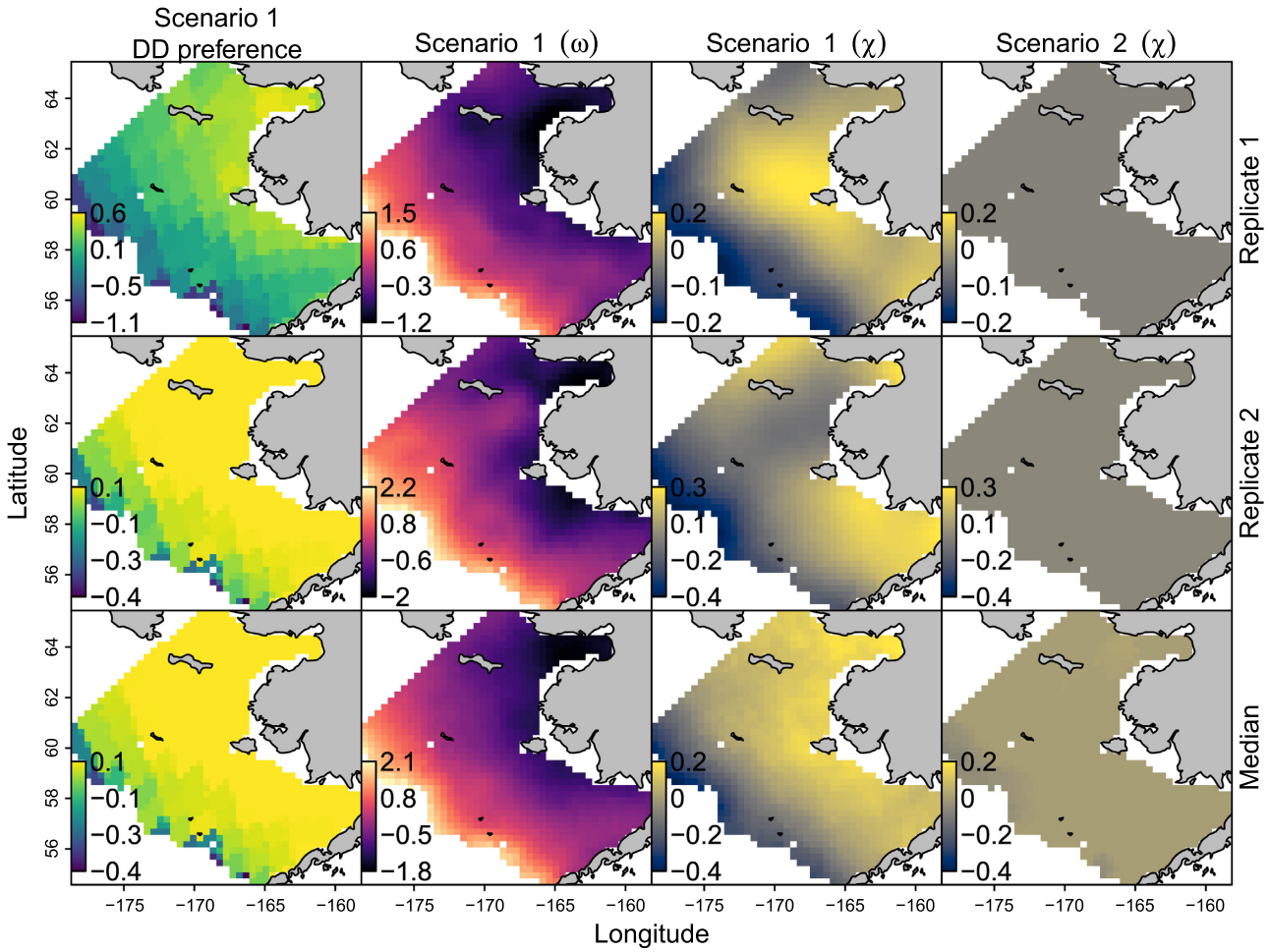


**Figure 3.** Component of preference due to density dependence averaged across all simulation years (left column, i.e. time-average of middle column from Figure 2), the estimated spatial component $\omega(s)$ (second column) and density-dependent response $\chi(s)$ (third column) for the "intermediate complexity" density-dependent SDM fitted to data simulated by the RAD simulation model given a density-dependent response for habitat preferences, as well as the estimated density-dependent response $\chi(s)$ (fourth column) when fitted to data simulated without density-dependent habitat selection (where the solid grey response indicates a median response of vanishingly close to 0). Note that the first and third columns are expected to show similar patterns (despite the large differences in model structure and resolution between simulation and estimation models), and that the third and fourth columns use the same colour-scale for each row (to highlight differences between these two scenarios).

mirrors the same pattern in the density-dependent response estimated by the SDM (Figure 3, third column). By contrast, the estimated density-dependent response typically goes to zero in the density-independent scenario (Figure 3, right column), indicating that the data are often sufficient to identify a small or negligible density-dependent response in this scenario.

Similarly, the majority of replicates in the density-independent scenario estimate a standard deviation approaching zero for this spatially varying response (Figure 4, grey histogram in right panel), while almost all replicates for the density-dependent scenario estimate a spatially varying term with $SD > 0.1$ (Figure 4, blue histogram in right panel). In all cases for both scenarios, the correlation between the spatial term, $\omega(s)$, and the density-dependent response, $\chi(s)$, is estimated to be negative, and it is substantially more negative in the density-dependent scenario than otherwise (Figure 4, left panel). Finally, the average of the density-dependent preference in the simulation model generally has a substantial (0.1–0.7) correlation with the estimated density-dependent response in the SDM (Figure 4, middle panel).

Inspecting results from the case-study application, fits of both the "two-stage" and "intermediate complexity" approaches are similar for all four species in the eastern Bering Sea. For arrowtooth flounder (Figure 5, top row), as expected, the density-dependent model estimates that increased abundance is associated with faster-than-proportional increases in density in the middle domain. Similarly, capelin shows a positive density-dependent response in the middle domain near the Pribiloff Islands (Figure 5, third row); unlike arrowtooth, it has higher expected density in the inner domain, such that an increase in abundance is associated with densities spilling offshore. Both tanner crab (Figure 5, second row) and Pacific cod (Figure 5, fourth row) show more localized responses to increasing density, with relative density for tanner increasing within hotspots in the northern middle domain, and Pacific cod increasing nearshore when abundance is high. In all cases, the index of median abundance $\beta(t) + \bar{\varepsilon}(t)$ used as covariate by the "intermediate complexity" model (black line in Figure 5, right column) is highly correlated with the estimate of log-abundance used by the "two-stage" approach (blue line in Figure 5, right column), but we

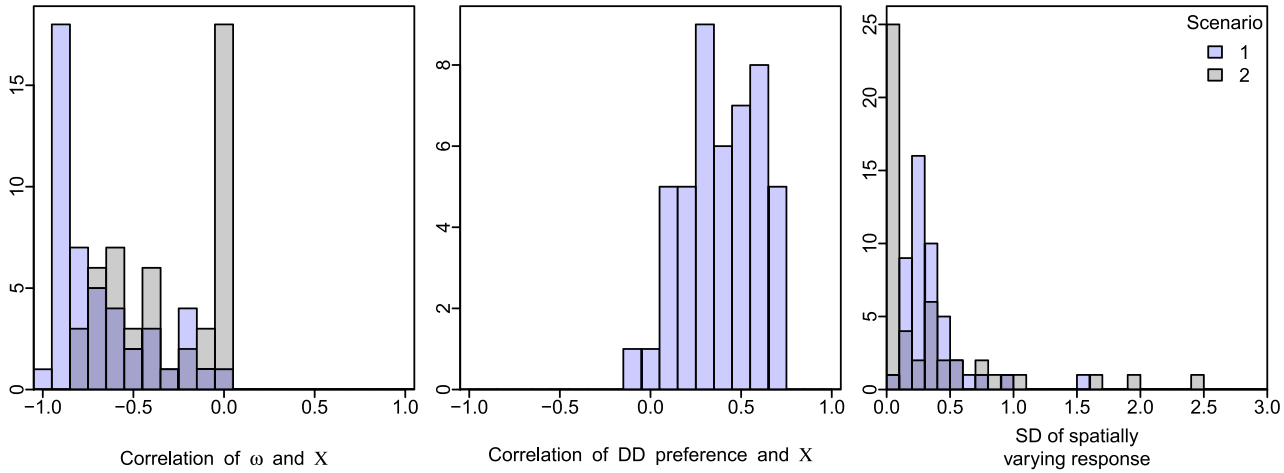


**Figure 4.** Correlation between spatial component $\omega$ and density-dependent response $\chi$ using the "intermediate complexity" density-dependent SDM (i.e. the second and third columns of Figure 3) for 50 replicates of the simulation scenarios with (Scenario 1 in blue) or without (Scenario 2 in grey) density-dependent habitat selection (left panel), the correlation between density-dependent response $\chi$ and the average across years of the component of preference due to density-dependence (middle panel, i.e. correlation between the first and third columns of Figure 3), as well as the estimated standard deviation of the spatially varying response $\chi$ (right panel).

note that the scale differs in particular for arrowtooth, and this results in turn in different scales (but similar spatial patterns) for the density-dependent response between the two approaches. Finally, projected density in the absence of residual spatio-temporal variation using the "intermediate complexity" approach isolates a density-dependent range shift for all four species (Figure 6). Both arrowtooth (Figure 6, top row) and capelin (Figure 6, third row) shows strong density-dependent range expansion, where the former expands inshore and the latter offshore given high abundance (e.g. contrasting range edge in Figure 6, right column). Similarly, tanner crab (Figure 6, second row) shifts northward in the middle domain and Pacific cod (Figure 6, fourth row) has greater density in nearshore habitat given high abundance.

## Discussion

We have introduced three statistical approaches for estimating density dependence within SDMs that do not require an auxiliary time-series of stock abundance, while exploring in detail an "intermediate complexity" approach using simulated and real-world data. These approaches extend previous spatially varying coefficient (SVC) models that used an independent time-series of abundance to model density dependence (Bacheler *et al.*, 2009, 2012; Bartolino *et al.*, 2011; Ciannelli *et al.*, 2012), and this study is the first (to our knowledge) to use a simulation experiment to test the performance of an SVC model to represent density dependence. In the reaction-advection-diffusion (RAD) simulation, the density-dependent SDM typically estimated a substantial response when habitat preferences were density-dependent, and a weak or nonexistent response when habitat preferences were density-independent. The estimated density-dependent response was correlated with the time-averaged component of preference due to density dependence, despite substantial differences in model structure between the explicit-movement simulation model and the regression-based estimation model. The case-study performed as expected for arrowtooth flounder, which has spilled into the middle domain during several decades of population increase (Bartolino *et al.*, 2011; Thorson *et al.*, 2016), and also showed previously undocumented responses for other Bering Sea species. Using the fitted models, we show that the density-dependent effect can be isolated to visualize density-dependent range shifts for each case-study species. We first discuss properties of the "intermediate complexity" approach, and then discuss three potential uses for this density-dependent SDM, including (1) exploring improvements in performance when informing spatial management procedures; (2) defining core habitat utilization; and (3) cross-testing between mechanistic and correlative SDMs. We then conclude by outlining a few avenues to extend the density-dependent SDM model.

The "intermediate complexity" approach (and the GAM approach that it builds upon) include several interesting and noteworthy properties:

1. Areas with a highly negative estimate of the spatially varying response (i.e. $\chi(s) < -1$) will show a decrease in density as total abundance increases (and an increase in density with decreasing abundance). We see responses this negative for all case-study species, and note that such behaviour cannot be explained solely by density-dependent habitat selection. However, it could arise in nature from nonlocal processes, e.g. seasonal movement combined with prey depletion could cause increases in total abundance to result in greater than proportional movement away from habitats with depleted forage.
2. The relative importance of two habitats may be inverted during changes in total abundance. For example, consider two sites $s_1$ and $s_2$. If $\omega(s_1) > \omega(s_2)$ and $\chi(s_1) < \chi(s_2)$, then density will be greater at $s_1$ than $s_2$ at sufficiently low abundance but the opposite will occur at sufficiently high abundance. Again, this property will not arise under density-dependent habitat selection in isolation, but may arise given the complexities of real-world spatial dynamics.
3. The area–abundance curve (i.e. a plot of area-occupied and total abundance) may be U- or dome-shaped. This can occur e.g. when average spatial distribution $\omega$ is highly different from density-dependent response $\chi$. In this case, periods with average abundance will have distribution concentrated in areas with

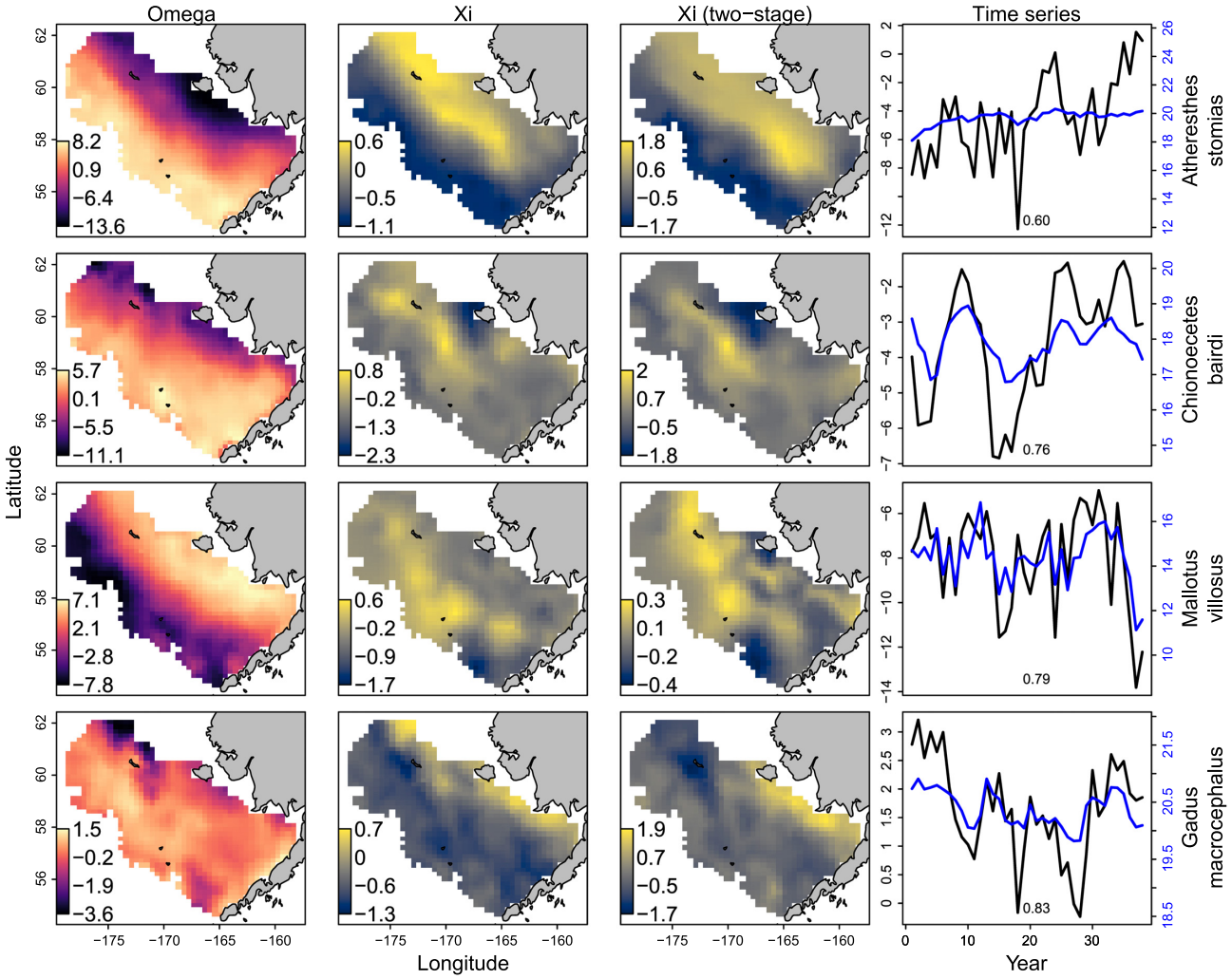


**Figure 5.** Case study results for the density-dependent SDM fitted to four species (rows from top to bottom: arrowtooth, tanner crab, capelin, and flathead sole) showing the average spatial component $\omega(s)$ and density-dependent response $\chi(s)$ for the "intermediate complexity" approach (first and second columns), the density-dependent response for the "two-stage" approach (third column), and a comparison of log-total biomass (blue line with right-hand *y*-axis labels) compared with the median log-density $\beta(t) + \bar{\varepsilon}(t)$ (black line with left-hand *y*-axis labels in fourth column).

high $\omega$, while periods with high abundance will have distribution concentrated in areas with high $\chi$, but intermediate abundance may have more homogenous distribution between these two extremes.

We note that all three properties occur in previous GAM models for density dependence, and these flexible behaviours may in some cases be useful for describing the complicated and context-dependent outcomes of density dependence observed in real-world populations (Avgar *et al.*, 2020).

Previous GAM approaches and our present approach differ most greatly in how they represent the time-series of "abundance" used to interpret density dependence. Previous GAM approaches did not require or enforce any similarity in the time-series of biomass used as covariate, relative to the estimate of total abundance generated by summing the GAM predictions of density across space. Therefore, the time-series used as covariate might show large differences in trend relative to the trend in the SDM; in these cases, the "density-dependent response" is a response to different dynamics than the abundance being estimated within the model. By contrast, all three approaches developed here attempt to ensure that the estimated trend in abundance (whether total abundance in the simple or Lagrange multiplier-approaches, or median abundance in the "intermediate complexity" approach) also matches the covariate that is used to drive density dependence. We see the tight link between SDM estimates of abundance and the resulting density-dependent response as a benefit of our proposed approach, in particular because most stocks in the US and worldwide do not have a stock assessment to generate an independent time-series of abundance for use within a density-dependent SDM (Neubauer *et al.*, 2018).

Range collapse resulting from declining abundance and density-dependent habitat selection has been implicated in past fishery collapses (Hutchings and Myers, 1994), and the density-dependent SDM provides one avenue to forecast range expansion/contraction given future changes in total abundance. Measurements of area occupied could be used to develop novel management strategies, including rotational closures as used for New England sea scallops

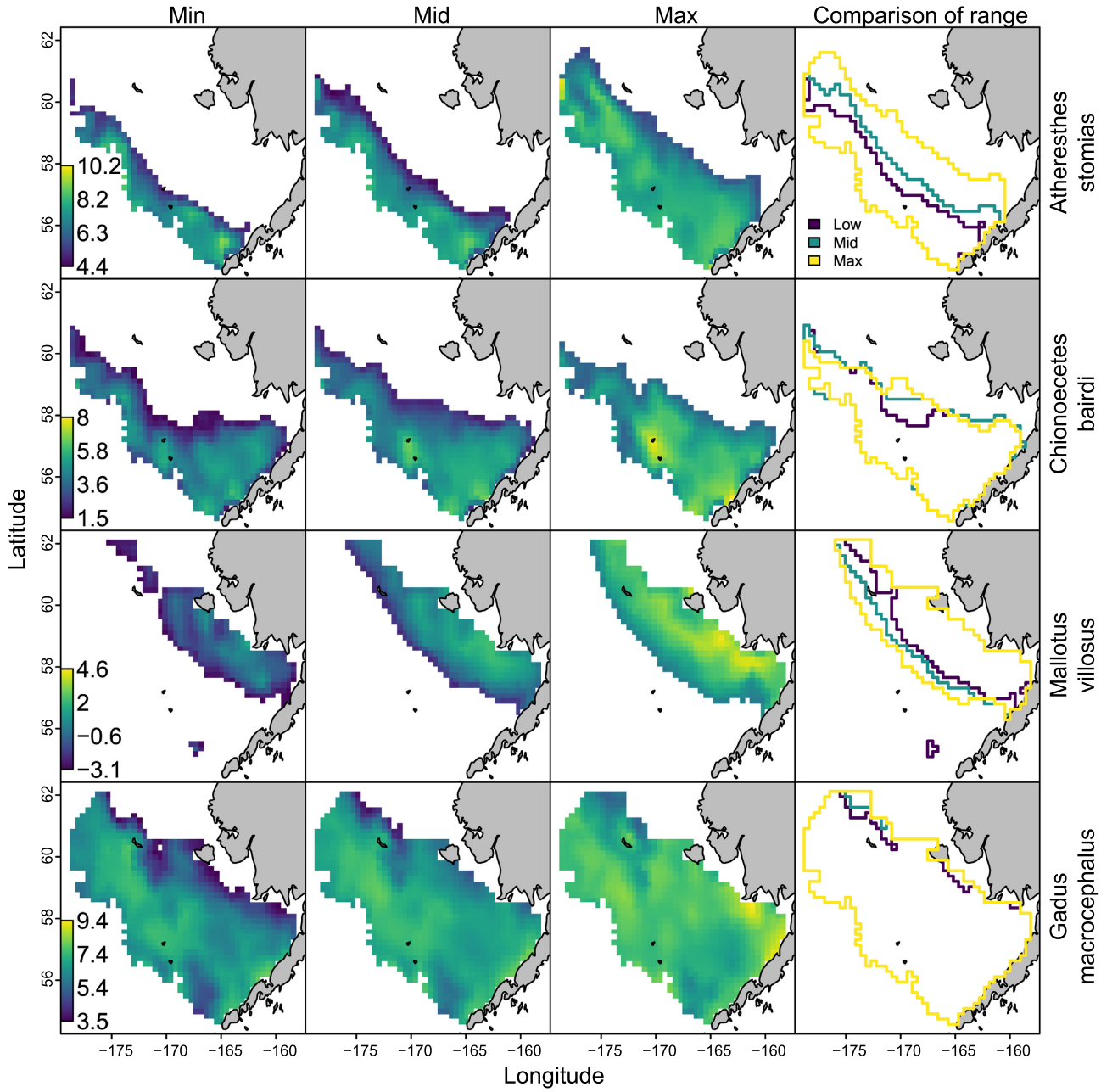


**Figure 6.** Illustration of projected densities $d(s,\ t)$ for four modelled species (rows) given the lowest (first column), intermediate (second column), or highest (third column) of the estimated range for intercept $\beta_t$ for the modelled years 1982–2019 in our case study, calculated when setting spatio-temporal term $\varepsilon\ (s,\ t) =\ 0$ for all locations and years to isolate the predicted effect of density dependence in forecasting distribution shifts. Note that densities less than 1% of the maximum value for that species and year are not plotted (so that white-space indicates the range edge based on this definition), but the colour legend is the same for all columns for a given species (to facilitate comparison of densities across columns), and the range edges are also overlaid (fourth column to allow comparison of areas gained/lost during an increase in median abundance).

(Gedamke *et al.*, 2004) or creation of new urban habitats for imperiled western monarch populations (Crone and Schultz, 2021), but these typically involve forecasted changes in which habitats are occupied under future population sizes. We, therefore, recommend testing our ability to forecast spatial dynamics using density-dependent vs. -independent SDMs. Most importantly, we recommend that future studies explore the relative importance of density-dependent and -independent responses using both within-sample inference (i.e. which mechanism causes a greater reduction in residual spatio-temporal variance and percent deviance explained) and retrospective skill testing (i.e. which mechanism can better improve forecasts of future dynamics). In particular, retrospective skill testing can be designed to measure performance for a specific management task, e.g. when used to inform a specific variable within a given management context (DeFilippo *et al.*, 2021). However, this will require extending (and further testing) the current approach to

include a density-dependent response resulting from the effect of environmental covariates.

We also believe that the density-dependent SDM has potential to provide fundamental insights into habitat preference, selection, and conservation. In particular, as populations decline in abundance (whether due to climate impacts or overfishing), they will sometimes retract to core habitats (MacCall, 1990). When doing so, they become particular sensitive to habitat disturbance and/or restoration within those core areas. We, therefore, propose that predictions of habitat utilization at the limit reference point for overfished status (e.g. $B_t < 0.2B_0$) is a useful metric of core habitat. These habitats could then warrant particular focus under US federal management, either as essential fishing habitat (EFH) or habitat areas of particular concern (HAPC). Alternatively, estimating a density-dependent response for a pair of species can show whether they are expected to have increased or decreased overlap given population increases. Similarly, a joint SDM involving a predator and prey could include an explicit link where the prey colonizes or avoids some habitats as predator abundance increases or decreases.

Finally, we note that the density-dependent SDM developed as an estimation model is an approximation to the RAD model used here as a simulation model. Specifically, the RAD simulation model includes an explicit habitat-preference function that can be informed via experimental lab and field studies, and could account for alternative forms of density dependence (i.e. habitat preference explicitly affected by local vs. regional densities). By contrast, the density-dependent SDM instead approximates this as an additive component within a regression, and therefore, cannot explicitly estimate a habitat-preference function. Importantly, we hypothesize that the SDM approximation is likely to have parameters that change more rapidly over time (i.e. are less stationary) than a more mechanistic model structure such as the RAD model. Recent statistical developments allow fitting the more mechanistic RAD model directly to fishery and survey data (Thorson *et al.*, 2021a), and we therefore recommend parallel development of the RAD and density-dependent SDM models. As shown here, the parallel development of these two approaches will at a minimum allow cross-testing, and could support a future study exploring which is likely to be a more stationary approximation to movement dynamics under climate change. It also remains unclear how easily each method can be extended to include new processes, e.g. including a density-dependent change in preference for habitat attributes, which shows up within a correlative SDM as an interaction between density and other variables. In these and other cases, we see a benefit to ongoing, parallel development of mechanistic and correlative SDMs.

## Data availability statement

The bottom trawl data used here are publicly available online as curated by the Alaska Fisheries Science Center (https://apps-afsc.fisheries.noaa.gov/RACE/groundfish/survey_data/data.htm).

## Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

## Acknowledgements

We thank the many scientists and volunteers who have supported the eastern Bering Sea bottom trawl survey used here. We also thank M. Goodman, L. DeFillipo, two anonymous reviewers, and V. Bartolino for helpful comments on a prior draft.

## References

Avgar, T., Betini, G. S., and Fryxell, J. M. 2020. Habitat selection patterns are density dependent under the ideal free distribution. Journal of Animal Ecology, 89: 2777–2787.

Bacheler, N. M., Bailey, K. M., Ciannelli, L., Bartolino, V., and Chan, K.-S. 2009. Density-dependent, landscape, and climate effects on spawning distribution of walleye pollock theragra chalcogramma. Marine Ecology Progress Series, 391: 1–12.

Bacheler, N. M., Ciannelli, L., Bailey, K. M., and Bartolino, V. 2012. Do walleye pollock exhibit flexibility in where or when they spawn based on variability in water temperature? Deep Sea Research Part II Topical Studies in Oceanography, 65–70: 208–216.

Barnett, L. A. K., Ward, E. J., and Anderson, S. C. 2021. Improving estimates of species distribution change by incorporating local trends. Ecography, 44: 427–439.

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

Brooks, E. N., and Deroba, J. J. 2015. When "data" are not data: the pitfalls of post hoc analyses that use stock assessment model output. Canadian Journal of Fisheries and Aquatic Sciences, 72: 634–641.

Ciannelli, L., Bartolino, V., and Chan, K.-S. 2012. Non-additive and non-stationary properties in the spatial distribution of a large marine fish population. Proceedings of the Royal Society B Biological Sciences, 279: 3635–3642.

Crone, E. E., and Schultz, C. B. 2021. Resilience or catastrophe? A possible state change for monarch butterflies in western North America. Ecology Letters 24: 1533–1538.

DeFilippo, L. B., Buehrens, T. W., Scheuerell, M., Kendall, N. W., and Schindler, D. E. 2021. Improving short-term recruitment forecasts for coho salmon using a spatiotemporal integrated population model. Fisheries Research, 242: 106014.

Dickey-Collas, M., Payne, M. R., Trenkel, V. M., and Nash, R. D. M. 2014. Hazard warning: model misuse ahead. ICES Journal of Marine Science, 71: 2300–2306.

Elith, J., Kearney, M., and Phillips, S. 2010. The art of modelling range-shifting species. Methods in Ecology and Evolution, 1: 330–342.

Foster, S. D., and Bravington, M. V. 2013. A Poisson–Gamma model for analysis of ecological non-negative continuous data. Environmental and Ecological Statistics, 20: 533–552.

Gedamke, T., DuPaul, W. D., and Hoenig, J. M. 2004. A spatially explicit open-ocean delury analysis to estimate gear efficiency in the dredge fishery for sea scallop *Placopecten magellanicus*. North American Journal of Fisheries Management, 24: 335–351.

Hanks, E. M., Hooten, M. B., and Alldredge, M. W. 2015. Continuous-time discrete-space models for animal movement. The Annals of Applied Statistics, 9: 145–165.

Hastie, T., and Tibshirani, R. 1993. Varying-coefficient models. Journal of the Royal Statistical Society Series B (Methodological), 55: 757–779.

Hutchings, J. A., and Myers, R. A. 1994. What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, of Newfoundland and Labrador. Canadian Journal of Fisheries and Aquatic Sciences, 51: 2126–2146.

Kearney, K., Hermann, A., Cheng, W., Ortiz, I., and Aydin, K. 2020. A coupled pelagic–benthic–sympagic biogeochemical model for the bering sea: documentation and validation of the BESTNPZ model (v2019.08.23) within a high-resolution regional ocean model. Geoscientific Model Development, 13: 597–650.

Keith, D. A., Rodríguez, J. P., Rodríguez-Clark, K. M., Nicholson, E., Aapala, K., Alonso, A., Asmussen, M. *et al.* 2013. Scientific foundations for an IUCN red list of ecosystems. Plos ONE, 8: e62111.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. 2016. TMB: automatic differentiation and laplace approximation. Journal of Statistical Software, 70: 1–21.

Lauth, R. R., and Conner, J. 2016. Results of the 2013 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate resources. NOAA Technical Memorandum, NMFS-AFSC-331. Alaska Fisheries Science Center, Seattle, WA.

Lindgren, F., ,Rue, H. and Lindström, J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B (Statistical Methodology), 73: 423–498.

Litzow, M. A., Ciannelli, L., Puerta, P., Wettstein, J. J., Rykaczewski, R. R., and Opiekun, M. 2018. Non-stationary climate–salmon relationships in the Gulf of Alaska. Proceedings of the Royal Society B Biological Sciences, 285: 20181855.

MacCall, A. D. 1990. Dynamic Geography of Marine Fish Populations. University of Washington Press, Seattle, WA.

Neubauer, P., Thorson, J. T., Melnychuk, M. C., Methot, R., and Blackhart, K. 2018. Drivers and rates of stock assessments in the united states. Plos ONE, 13: e0196483.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology, Evolution, and Systematics, 37: 637–669.

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. 2013. Marine taxa track local climate velocities. Science, 341: 1239–1242.

Planque, B., Loots, C., Petitgas, P., Lindstrøm, U. L. F., and Vaz, S. 2011. Understanding what controls the spatial distribution of fish populations using a multi-model approach. Fisheries Oceanography, 20: 1–17.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/, Last access date: November 20, 2020.

Shepherd, T. D., and Litvak, M. K. 2004. Density-dependent habitat selection and the ideal free distribution in marine fish spatial dynamics: considerations and cautions. Fish and Fisheries, 5: 141–152.

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

Thorson, J. T., and Kristensen, K. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fisheries Research, 175: 66–74.

Thorson, J. T., Rindorf, A., Gao, J., Hanselman, D. H., and Winker, H. 2016. Density-dependent changes in effective area occupied for sea-bottom-associated marine fishes. Proceedings of the Royal Society B Biological Sciences, 283: 20161853.

Thorson, J. T., and Barnett, L. A. K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science, 74: 1311–1321.

Thorson, J. T. 2019. Measuring the impact of oceanographic indices on species distribution shifts: the spatially varying effect of cold-pool extent in the eastern Bering Sea. Limnology and Oceanography, 64: 2632–2645.

Thorson, J. T. 2019. Guidance for decisions using the vector autoregressive spatio-temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research, 210: 143–161.

Thorson, J. T., Ciannelli, L., and Litzow, M. A. 2020. Defining indices of ecosystem variability using biological samples of fish communities: a generalization of empirical orthogonal functions. Progress in Oceanography, 181: 102244.

Thorson, J. T., Barbeaux, S. J., Goethel, D. R., Kearney, K. A., Laman, E. A., Nielsen, J. K., Siskey, M. R. *et al.* 2021. Estimating fine-scale movement rates and habitat preferences using multiple data sources. Fish and Fisheries,22, 1359–1376. (last accessed 18 October 2021).

Thorson, J. T., Hermann, A. J., Siwicke, K., and Zimmermann, M. 2021. Grand challenge for habitat science: stage-structured responses, nonlocal drivers, and mechanistic associations among habitat variables affecting fishery productivity. ICES Journal of Marine Science.78, 1956–1968,

Walters, C., and Maguire, J.-J. 1996. Lessons for stock assessment from the northern cod collapse. Reviews in Fish Biology and Fisheries, 6: 125–137.

Warton, D. I., and Shepherd, L. C. 2010. Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. The Annals of Applied Statistics, 4: 1383–1402.

Zimmermann, M., and Benson, J. L. 2013. Smooth sheets: how to work with them in a GIS to derive bathymetry, features and substrates. NOAA Technical Memorandum, NMFS-AFSC-249. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center.

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

*Handling Editor: Valerio Bartolino*