Methods in Ecology and Evolution British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 990–1002 doi: 10.1111/2041-210X.12567

Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity

James T. Thorson[1]*, Malin L. Pinsky[2†] and Eric J. Ward[3†]

[1]*Fisheries Resource Assessment and Monitoring Division (FRAM), Northwest Fisheries Science Center, National Marine Fisheries Service (NMFS), NOAA, 2725 Montlake Blvd. E, Seattle, WA 98112, USA;* [2]*Department of Ecology, Evolution, and Natural Resources, Rutgers University, 14 College Farm Road, New Brunswick, NJ 08901, USA; and* [3]*Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service (NMFS), NOAA, 2725 Montlake Blvd. E, Seattle, WA 98112, USA*

Summary

1. Changing climate is already impacting the spatial distribution of many taxa, including bees, plants, birds, butterflies and fishes. A common goal is to detect range shifts in response to climate change, including changes in the centre of the population's distribution (the centre of gravity, COG), population boundaries and area occupied. Conventional estimators, such as the abundance-weighted average (AWA) estimator for COG, confound range shifts with changes in the spatial distribution of available survey data and may be biased when the distribution of survey data shifts over time. AWA also does not estimate the standard error of COG in individual years and cannot incorporate data from multiple survey designs.

2. To explicitly account for changes in the spatial distribution of survey effort, we propose an alternative species distribution function (SDF) estimator. The SDF approach involves calculating distribution metrics, including COG, population boundary and area occupied, directly from the predicted species distribution or density function. We illustrate the SDF approach using a spatiotemporal model that is available as an R package. Using simulated data, we confirm that the SDF substantially decreases bias in COG estimates relative to the AWA estimator. We then illustrate the method by analysing data from two data sets spanning 1977–2013 for 18 marine fishes along the U.S. West Coast.

3. In our case study, the SDF estimator shows significant northward shifts for six of 18 species (with southward shifts for only 2), where two species (darkblotched and greenstriped rockfishes) have both a northward shift and a decreased area occupied. Pelagic species (e.g. Pacific hake and spiny dogfish) have more variable distribution than bottom-associated species. We also find substantial differences between AWA and SDF estimates of COG that are likely caused by shifts in sampling distribution (which affect the AWA but not the SDF estimator).

4. We caution that common estimators for range shift can yield inappropriate inference whenever sampling designs have shifted over time. We conclude by suggesting further improvements in model-based approaches to analysing climate impacts, including methods addressing the impact of local and regional temperature changes on species distribution.

Key-words: abundance-weighted average, California Current, centre of gravity, climate change, range shifts, spatiotemporal model, species distribution model

Introduction

Climate change and variability have already impacted the spatial distribution of many different taxa including butterflies (Parmesan *et al.* 1999), plants (Kelly & Goulden 2008), bees (Kerr *et al.* 2015), birds (Hitch & Leberg 2007) and marine fishes (Pinsky *et al.* 2013). Climate impacts have often been identified via changes in spatial distribution, due to the wide availability of spatially referenced survey data for many taxa as well as the relative ease of detecting changes in population distribution over time. Changes in distribution can have immediate relevance to local economies and international diplomacy (Cheung *et al.* 2012), as well as species interactions and evolutionary trajectories (Garroway *et al.* 2010). Climate change has also been documented to affect many biological processes, including the allocation of energy between growth and reproduction, community composition and phenology (Parmesan 2006).

Estimating directional changes in a spatial distribution is not unique to population biology – similar problems exist in quantifying distributional changes in sea ice, temperature, rainfall and patterns of disease in response to climate change. In all of these problems, it can be difficult to distinguish between long-term (multidecadal) trends caused by anthropogenic climate change, interdecadal variability and

---

*Correspondence author. E-mail: .James.Thorson@noaa.gov

[†]MLP and EJW contributed equally.

short-term fluctuations in spatial distribution. As a result, identifying trends in spatial distribution due to climate change generally requires analysing data from decadal or longer timescales (Hitch & Leberg 2007). Long time series regarding species distribution are rare (Parmesan *et al*. 2005), and much could be learned if shorter time series could be stitched together. However, combining data from multiple sampling protocols poses serious difficulties, including quantifying the relative performance of survey efficiency among different survey teams or sampling protocols (Sauer, Peterjohn & Link 1994), accounting for spatial differences in sampling intensity and estimating the uncertainty in any resulting estimate of distribution. Even within a long-running survey, such as the North American Breeding Bird Survey, volunteer skill and training, sampling design, and sampling effort may vary regionally and over time. For example, sampling sites may be added or removed in any given year, sampling effort may go up or down at particular sites, or the sample design itself may include random site selection (e.g. a stratified random design). Authors have tried to deal with these complications by trimming survey data to only consistently surveyed sites and time periods with consistent survey methods (Pinsky *et al*. 2013) or by stratifying the survey area (Rindorf & Lewy 2006; Dulvy *et al*. 2008; Engelhard, Righton & Pinnegar 2014). The former is often unsatisfactory because it requires discarding data. The latter estimator will eliminate the bias arising from changes in survey sampling intensity given that each stratum is sufficiently small, but will also lose precision in this case unless the average within each stratum is calculated using a shrinkage estimator (i.e. by borrowing information from nearby strata).

In this study, we compare conventional approaches to quantifying changes in species distribution to a new model-based estimator. We first show that the conventional 'abundance-weighted average (AWA)' estimator for the centre of a species distribution (termed the 'centre of gravity', COG) confounds true changes in species distribution with changes in the distribution of sampling effort. The conventional approach also does not estimate a standard error for distribution metrics and therefore cannot distinguish between sampling variation and significant interannual variability. We then show how to estimate COG, area occupied and population range using any model that estimates a species distribution function (SDF). SDFs are estimated by a wide class of models including, for example occupancy models and presence-only models. The SDF approach to estimating distribution shifts can account for shifts in sampling effort over time (or concentrations of effort in particular combinations of years and areas), separating the observation process from the true underlying spatial distribution, and hence account for changes in survey methods or locations when combining data from multiple sampling programmes. The SDF approach also estimates standard errors that can be used to distinguish whether interannual variation in distribution is biologically important or an artefact of limited sampling data. In this study, we apply the SDF estimator to data for 18 bottom-associated fishes off the U.S. West Coast, integrating data across two survey programmes (1977–2004 and 2003–2013). Using the SDF model, this case study illustrates that six species show a statistically significant, northward shift in distribution from 1977 to 2013, while only two show a significant southward shift. All code for replicating this analysis (as well as example data for surveys of marine fishes in the California Current, Gulf of Alaska, Eastern Bering Sea and South Africa) is publicly available (https://github.com/nwfsc-assess/geostatistical_delta-GLMM/wiki).

Materials and methods

A wide range of metrics have been used or proposed for understanding shifts in species distributions, including the COG, location of population boundaries, the area occupied, the mean temperature of occurrence or the thermal preferences of species in a local community (Perry *et al*. 2005; Devictor *et al*. 2008). Below, we describe some of the conventional estimators as well as a novel SDF approach from which a wide range of metrics can be estimated.

CONVENTIONAL ESTIMATORS FOR SHIFTS IN DISTRIBUTION

Conventional estimators use available data on the location of species occurrences (and often abundance as well), often processed so as to attempt to ensure comparability through time. For example, the AWA estimator has been widely used to track changes in the COG of a population's distribution (Perry *et al*. 2005; Kelly & Goulden 2008; Pinsky *et al*. 2013; Hiddink, Burrows & García Molinos 2015). This estimator calculates the COG as the average location of samples in a given year, where each location is weighted by the abundance (measured either as numbers or biomass) of the species encountered at that location. This estimator can be applied in multiple dimensions (e.g. latitude and longitude), but as one example the latitudinal COG, $\overline{\text{Latitude}}$, is as follows:

$$\overline{\text{Latitude}} = \frac{\sum_{i=1}^{n} c_i \times \text{Latitude}_i}{\sum_{i=1}^{n} c_i}, \qquad \text{eqn 1}$$

where n is the number of available samples, c_i is the abundance of the species in the ith sample, and Latitude_i is the latitude of the ith sample. This estimator confounds changes in the spatial distribution of sampling with range shifts for sampled species (see Appendix S1, Supporting information) and therefore is only suitable when the distribution of sampling is constant over time. Conventional methods for calculating standard errors (e.g. sample variance computations or non-parametric bootstrapping) are not appropriate for this estimator because residuals are likely autocorrelated. The inability to estimate standard errors complicates efforts to distinguish between random variation in the estimator and either interannual or multiyear trends in COG. Inferring multiyear trends from the AWA estimator is generally done by estimating the COG in each time period, then fitting a linear regression model to estimate temporal trends. Significant trends in COG are identified using a Wald test given the null hypothesis that the trend coefficient is equal to zero.

The same data can also be used to estimate other metrics for changes in distribution. While population boundaries are difficult to observe because species detectability can decline as a species become rarer, estimators for population boundaries include the absolute highest and lowest latitudes at which a species is observed or an average of the three highest or lowest latitudes (Perry *et al*. 2005). Similarly, area occupied can be computed as the minimum number of grid cells that contain 90% of species abundance in each year (Fisher & Frank 2004). Similar to the AWA estimate of COG, these metrics lack a well-defined standard error. Trends over time are often detected with a linear regression model.

SPECIES DISTRIBUTION FUNCTION ESTIMATOR FOR SHIFTS IN DISTRIBUTION

As an alternative to the conventional sample-based methods, we propose a model-based estimator for analysing shifts in distribution. The model-based approach begins by defining a SDF d, where the value of this function $d(s)$ at any given location s describes the expected density of the population within a spatial domain $s \in \mathcal{D} \in \mathbb{R}^2$. Model parameters are estimated based on available data, and parameters and data are then used to predict the SDF $\hat{d}(s)$ at all locations s within a given spatial domain. We advocate estimating model parameters, predicting density $\hat{d}(s)$ and calculating summary statistics (e.g. $\overline{\text{Latitude}}$) within a single model, so that the standard error of the summary statistic represents estimation uncertainty for model parameters. Although our case study example includes variation over time and incorporates data among different surveys, the SDF approach could also be used in simpler applications (e.g. species distribution or occupancy models without any variation over time).

The predicted SDF $\hat{d}(s)$ can be used to calculate many possible statistics representing distribution shifts, for example the COG, population boundaries and area occupied by a given species (Appendix S2). For example, latitudinal COG can be estimated as:

$$\overline{\text{Latitude}} = \frac{\sum_{i=1}^{n_s} \hat{d}(s_i) \times \text{Latitude}\,(s_i)}{\sum_{i=1}^{n_s} \hat{d}(s_i)} \qquad \text{eqn 2}$$

where n_s is the number of locations (e.g. grid cells) used to summarize a spatial domain, $\hat{d}(s_i)$ is predicted density at location s_i, and Latitude (s_i) is latitude for this location. This calculation (eqn 2) resembles the AWA approach (eqn 1), except replacing a weighted average of sample locations with a weighted average of predictive locations. To calculate population boundaries, we calculate the cumulative distribution for abundance in northings and eastings, and then identify the 5th and 95th percentiles of this cumulative distribution. For area occupied, we first estimate a population kernel that summarizes the distribution of the species (Woillez, Rivoirard & Petitgas 2009) and then calculate the volume of this kernel.

This model-based estimator provides several benefits relative to conventional estimators like AWA:

1 It allows data from multiple sampling programmes to be used when estimating shifts in distribution, while controlling for changes in sampling efficiency and seasonal timing between surveys;
2 It estimates the standard error of COG estimates, which can be used to determine whether interannual variability and multiyear trends exceed the variability that is expected from random chance; and
3 It accounts for changes in the spatial distribution of sampling effort from 1 year to the next (whereas the AWA and other conventional estimators do not).

A model-based approach to estimating shifts in distribution has been used in previous ecological studies (Elith, Kearney & Phillips 2010; Matthews *et al.* 2011), although these have generally used habitat envelope modelling (i.e. predicting density as a function of measured habitat variables) rather than by estimating spatially correlated latent variables (as we do in our case study example). Previous research suggests that including spatial autocorrelation will improve predictions of species density compared with using only measured habitat variables (Bahn & McGill 2007; Shelton *et al.* 2014), so we hypothesize that incorporating spatial autocorrelation will also improve our estimates of range shifts relative to habitat envelope approaches.

APPLYING THE SDF ESTIMATOR IN A SPATIOTEMPORAL MODEL

For our case study application, we estimate the species density function $d(s, t)$ for multiple time intervals, $t \in \{1, 2, \ldots, T\}$ where T is the number of modelled intervals. We use a conventional ‘delta model’ [also known as a ‘hurdle’ model (Martin *et al.* 2005)] which decomposes density $d(s, t)$ into two components: (i) the probability $p(s, t)$ of encountering the species at a given location and time and (ii) the expected density $r(s, t)$ of the species when encountered:

$$d(s, t) = p(s, t)r(s, t).$$

We specify that each of these processes follows a first-order random walk process (Cressie & Wikle 2011), that is encounter probability $p(s, t)$ and expected density when encountered $r(s, t)$ in year t depend only upon their value the year before ($p(s, t-1)$ and $r(s, t-1)$) and the spatial distribution of the population is otherwise independent among years (see Thorson *et al.* 2015b and Appendix S3 for further details).

We next specify a distribution for sampling data, that is the probability that the ith survey sample is positive (i.e. that the species is encountered) is affected by encounter probability $p(s_i, t_i)$ at the location s_i and year t_i for sample i, as well as additional ‘catchability’ variables $z_p(i)$. Similarly, the survey sample given that the species was observed follows a log-normal distribution, where the expectation is affected by expected density $r(s_i, t_i)$ as well as catchability variables $z_r(i)$. Catchability variables are defined as those measurable factors that influence expected catch rates independently of changes in true density at the location of the sample. Density $d(s, t)$ is calculated for a reference value of these variables, $z_p(i) = z_r(i) = 0$. Therefore, any process that is included as a catchability variable is ‘controlled for’ when predicting density or associated metrics of range shift.

Parameters governing the stochastic process $p(s, t)$ and $r(s, t)$ are estimated using survey data from multiple survey designs. In the following, we assume that sampling in each year follows a probability sampling design, that is that there exists some stochastic process $\mathcal{P}$ governing the spatial allocation of samples, where $\mathcal{P}(s, t)$ is the probability that location s will be sampled in time t. We further assume that this sampling process is statistically independent of population density $d(s, t)$. This assumption is appropriate given that the sampling follows a pre-determined protocol which was designed independently from dynamics for any single species. However, the assumption would be invalid in other instances, for example if we used data from a fishery targeting a particular species. Given these two assumptions, we can estimate fixed effects by maximizing the marginal likelihood of the data without explicitly estimating the sampling process $\mathcal{P}$ during parameter estimation (Cressie *et al.* 2009). In the following, we treat as fixed the coefficients for measured variables, the magnitude of spatiotemporal variation, the spatial scale for encounter probability and positive densities, the catchability coefficients and the residual variance of positive catch rates (see Appendix S3 for further details regarding parameters). We treat the spatiotemporal variation as random effects and approximate the probability of these random effects using a stochastic partial differential equation approximation (Lindgren, Rue & Lindström 2011; Thorson *et al.* 2015c).

To estimate parameters, we maximize the marginal likelihood with respect to fixed effects. We approximate the marginal likelihood of fixed effects while integrating across random effects via the Laplace approximation using the Template Model Builder (TMB) software (Kristensen 2014; Kristensen *et al.* in press). Estimating the SDF model using TMB involves five steps: (i) we specify the joint likelihood of data and random effects in a template file; (ii) for a proposed set of fixed effects, TMB identifies values of random effects that maximizes the joint likelihood; (iii) given these values for fixed and random effects, TMB calculates the second derivatives of the joint likelihood with respect to random effects and calculates the Laplace approximation to the marginal likelihood; (iv) TMB calculates the gradient of the Laplace approximation to the marginal likelihood with respect to fixed effects; and (v) we use the marginal likelihood and its gradients via a conventional nonlinear minimizer in the R statistical environment (R Core Team 2014) to identify maximum-likelihood estimates of fixed effects. Standard errors for parameters and derived quantities (e.g. the COG) are then calculated via the inverse Hessian matrix and delta method.

CASE STUDY APPLICATION

As a case study, we analyse two different survey data sets spanning 1977–2013 in the west coast of the USA. These surveys are designed to monitor changes in the biomass of fish populations that are targeted by commercial and recreational fisheries of the U.S. West Coast. The first survey (subsequently called the 'triennial survey') was operated by the Alaska Fisheries Science Center every third year from 1977 to 2004 (for a total of 10 survey years), with an average of 484 samples per year (Weinberg *et al.* 2002). The second survey (subsequently called the 'annual survey') was operated by the Northwest Fisheries Science Center every year 2003–2013 (for a total of 11 survey years) and has on average 629 samples per year (Bradburn, Keller & Horness 2011). Triennial and annual surveys also use different vessels and sampling gear, and these differences are likely to cause differences in sampling efficiency between surveys. Pinsky *et al.* (2013) previously analysed range shifts using the AWA estimator applied to data from the triennial survey, but did not incorporate data from the annual survey due to concerns about changes in survey methods. We here demonstrate the implications of analysing data from both surveys using the conventional or SDF estimators. This demonstration serves two purposes:

1. It illustrates the potential differences in inference arising from conventional and SDF estimators when applied to data where the spatial distribution of sampling changes over time; and
2. It updates the Pinsky *et al.* (2013) analysis when incorporating data from an additional 11 years of data (i.e. the 2003–2013 period in the annual survey).

We restrict our analysis to tows that occur within the spatial domain of both surveys, and this decreases sample sizes available for each survey (triennial: 434; annual: 412 samples per year on average). We use this restriction for both SDF and AWA estimators (although it is not necessary for the SDF estimator) to ensure that results are comparable between methods. For the AWA analysis, we fit separate linear regressions to the triennial and the annual survey data, reflecting their substantially different locations and methods.

The spatial sampling intensity function differs between the two survey designs even within the restricted area that is within the spatial domain of both surveys, and spatial sampling intensity also varies among years within a given sampling design. The two surveys also differ in seasonal timing and likely differ in terms of sampling efficiency (a.k.a. catchability, i.e. the proportion of fish captured that are located within the sampled area). In the SDF analysis, we therefore include the following catchability variables:

1. Julian calendar date j_i for each sample; and
2. an indicator function v_i for survey (where $v_i = 1$ if the sample is from the triennial survey and zero otherwise)

such that catchability variable $z_p(i) = z_r(i) = (j_i, v_i)^T$ for sample *i*. These two variables (*j* and *v*) represent linear offsets of encounter probabilities (in logit space) and positive catch rates (in log space). Both variables have sufficient contrast to be estimated with a low standard error, that is Julian calendar date has broad overlap among surveys (Fig. 1) and both annual and triennial surveys were conducted in 2004. Including these catchability covariates, while not including them when predicting density for calculating COG, 'filters out' the variation that is attributable to calendar date or survey design, and species distributions and COG are then predicted for a standard date (July 31st) and the triennial survey. We do not include any variables *a priori* as predictors of spatial variation in density (i.e. $x(s)$ is not included in the model) and therefore account for all spatiotemporal variation via the random walk process ($\varepsilon_p(s, t)$ and $\varepsilon_r(s, t)$, see Appendix S3). While future studies could explore oceanographic or habitat predictors of species distribution using the SDF model (Shelton *et al.* 2014), we do not do so to ensure that results are strictly comparable to the AWA estimator (which does not typically incorporate predictor variables).

For computational reasons, we approximate spatiotemporal variation ($\varepsilon(s, t)$) as being piecewise constant at a fine spatial scale. We do this by estimating functions at $n_x = 500$ 'knots', which are selected via a *k*-means clustering algorithm applied to location of sampling data. Each location *s* within the survey domain then has function value

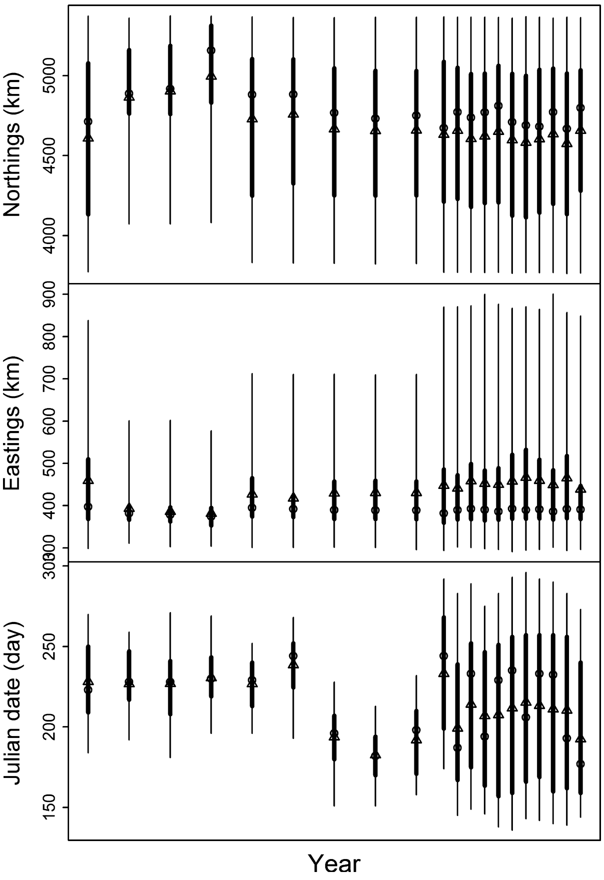

Fig. 1. Median (circle), mean (triangle), interquartile range (thick black line) and total range (thin black line) for the spatial location (top row: northings, middle row: eastings) and seasonal location (bottom row: Julian calendar date starting Jan. 1) in every year of West Coast sampling used in this study. The triennial survey (every third year 1977–2004) shows large differences in sampling distribution relative to the annual survey (every year 2003–2013), particularly with respect to northings and Julian date.

$\varepsilon(s_x, t)$ where s_x is the location of the knot x that is closest to location s. We confirm that results are qualitatively similar when increasing the number of knots (i.e. $n_x = 1000$). We then compare estimates of the change in distribution between AWA and SDF estimators to examine whether these two methods generate different inference regarding distribution shifts in this region.

VALIDATION WITH SIMULATED DATA

We also conduct a simulation experiment to illustrate the magnitude of bias that arises from using either the AWA or SDF estimators given the timing and location of samples that are available. In this experiment, we simulate population density $d(s, t) = p(s, t)r(s, t)$ in each of 15 979 grid cells (each is 2 × 2 nautical miles) that are included in the domain of the surveys in our analysis. We simulate density under four scenarios, representing (i) constant, (ii) highly variable, (iii) northward shifts or (iv) southward shifts in population COG (Appendix S4 for details). We then simulate sampling that occurs in the same years and at the same locations as in the annual and triennial surveys. We analyse the simulated data using the AWA and SDF estimators and compare the estimate of northward COG with its true value.

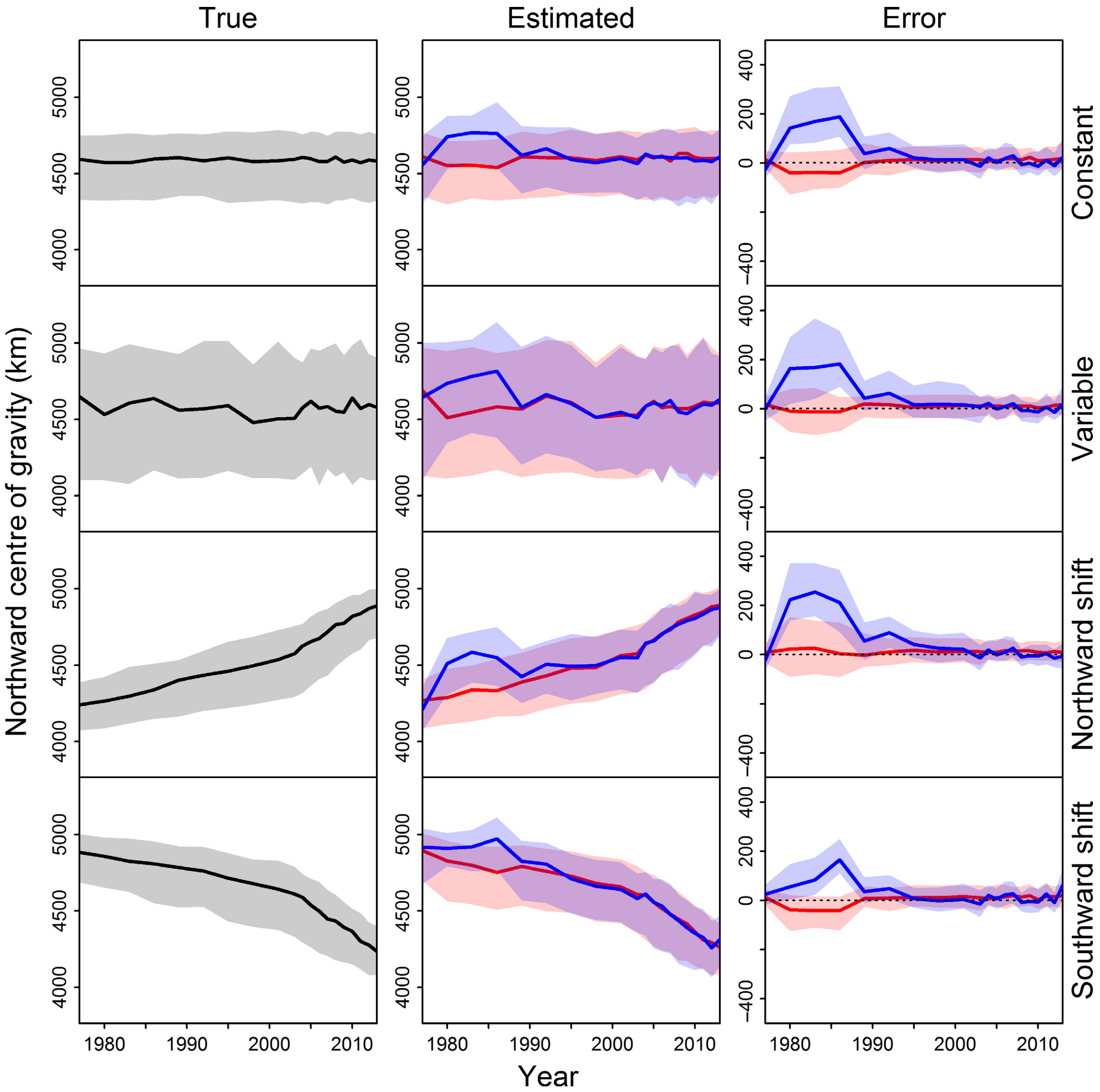

Fig. 2. Results from the simulation experiment for four treatments involving different shifts in species distribution over time (columns, See Validation with Simulated Data section of Materials and methods for descriptions), showing the northward centre of gravity for simulated abundance (left column; black line: median; shaded area: interval encompassing 80% of simulation replicates), the estimated centre of gravity for the species distribution function (SDF; red) and abundance-weighted average (AWA; blue) estimators (middle column) and the error (Est-True) for the SDF and AWA estimators (where a well-performing model will have error close to zero, as indicated by the dotted line).

Results

CHANGES IN THE DISTRIBUTION OF SAMPLING EFFORT FOR WEST COAST FISHES

Examination of the average latitude and longitude of survey samples in each year (Fig. 1) illustrates the magnitude of spatial changes in sampling intensity over time. In particular, the median latitude of samples was nearly 45°N in 1986, but was close to 40°N from 2005 to 2012. The distribution of the triennial survey varied greatly during the 1980s and did not sample far south of the San Francisco Bay in 1980, 1983 and 1986 (Appendix S5). Therefore, even a population that was uniformly distributed throughout this survey domain would appear to move south from the 1980s (during the early triennial sampling period) to the 2000s (during the annual sampling period).

SIMULATION EXPERIMENT

Results from our simulation confirm that the SDF estimator greatly mitigates the bias seen in the AWA estimator (Fig. 2). In particular, the AWA estimator is positively biased in all scenarios from 1977 to 2003 (i.e. during the triennial survey sampling period). This positive bias is greatest in years when the triennial survey had sampling distributed in more northerly areas (i.e. 1980, 1983, 1986). The bias is negligible for both estimators during the annual sampling period (2003–2013), and the SDF estimator has a small positive bias during early years for the ‘northward shift’ scenario and a small negative bias for the ‘southward shift’ scenario. Exploratory analysis indicates that the SDF estimator smooths density estimates for the unsampled, southward portion of the population domain for early years (1980, 1983 and 1986) towards the estimated distribution in 1989. Decreased bias using the SDF estimator is particularly evident for years 1989–2002, when the AWA is positively biased by nearly 50 km, while the SDF is approximately unbiased.

CASE STUDY

A plot of population density function $d(s, t)$ for Pacific hake (Fig. 3) is characteristic of the patterns estimated by the SDF model for West Coast species. For Pacific hake, the areas offshore of Point Conception and the San Francisco Bay generally have high density, but density near the San Francisco Bay is particularly high from 2001 to 2013, and areas offshore of southern Oregon have particularly high densities from 2003 to 2008. In aggregate, the population has a northward distribution during 1989–2009 and relatively southward distribution in 1977–1986 and 2010–2013 (Fig. 4). Changes among years often exceed the standard error of the estimated COG (e.g. 2009–2010), indicating changes that are statistically significant. A comparison of the model-based estimate of COG and the AWA (Fig. 4) again shows large differences between estimators in 1980–1986 (when sampling had greater intensity northwards in the California Current) and relatively small differences in 2003–2013 (when sampling intensity is close to uniform north of Point Conception).

We next present estimates of the COG for each of 18 species along the north–south axis (i.e. to detect northward or

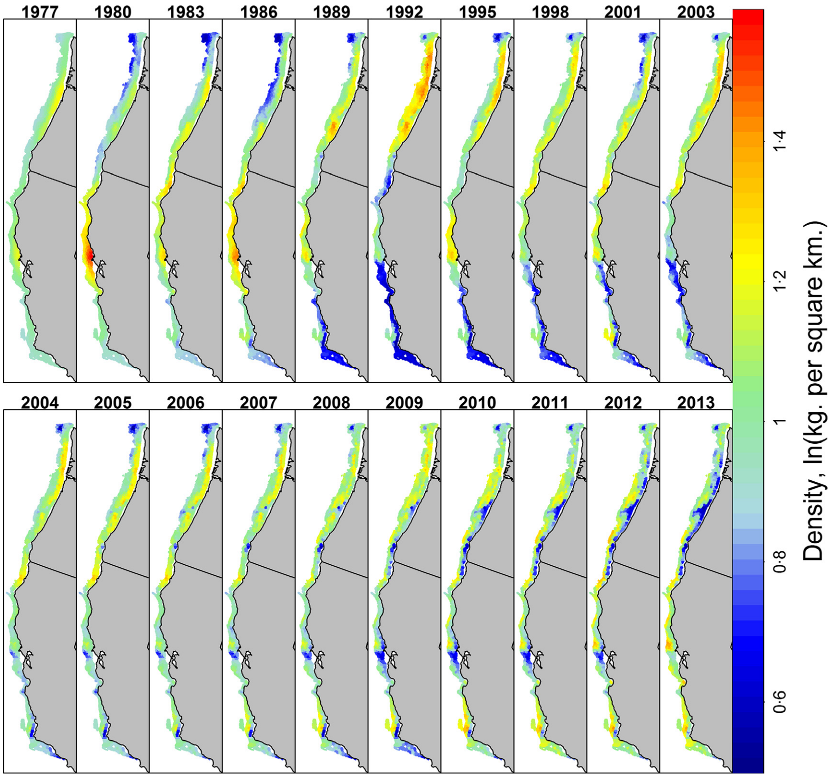

Fig. 3. Estimated species distribution function $d(s, t)$ for Pacific hake in every year rescaled for clarity of presentation to have a mean of one (where the coloured area represents the spatial domain that is included in this analysis).

southward movement over time, Fig. 5) for both AWA and model-based estimators (while analysing the triennial and annual surveys separately with the AWA estimator). The two estimators show different trends for many species during both the triennial and annual periods, for example Dover sole which has a northward trend during the triennial period (1977–2004) in the SDF estimator but a southward trend in AWA estimator. Inspecting results across species during the triennial period (1977–2004), the AWA estimator shows a shift that is more southerly than the SDF estimator for 16 of 18 species, where this southern shift is in-line with the southern trend in the location of sampling during triennial survey (Fig. 6). In particular, the AWA estimator shows significant southward shifts for two species and northward trends for one species from 1977 to 2004, while the SDF estimator shows significant northward shifts for seven species and significant southward shifts for only one species. Inspecting confidence intervals in the SDF estimator (Fig. 5) also illustrates that pelagic species (Pacific hake and spiny dogfish) generally have interannual variability that is large relative to confidence interval width, whereas rockfishes generally have less evidence of interannual variability in COG.

Inspecting results from the SDF estimator for each species and all available years (1977–2013, Fig. 7) illustrates six of the species have significant northward shifts, while only two (Pacific hake and sablefish) have significant southward shifts. Both hake and sablefish have had approximately steady area occupied during this period, so shifts in distribution have been caused by displacement of the population range, rather than expansion or contraction. Similarly, four of the six species with northward shifts have approximately steady area occupied. However, two species (darkblotched and greenstriped rockfishes) show both significant northward shifts and substantial decreases in area occupied (Fig. 8), indicating that the northward shift is caused by decreased density in southern habitats. This decreased density in southern habitats is also reflected in SDF estimates of the southern population boundary, which has moved northwards for both species (Appendix S6).

Discussion

Understanding the consequences of climate change on terrestrial and aquatic species requires unbiased estimates of how species' ranges shift. In this study, we have demonstrated that a SDF estimator for distribution shifts offers three benefits over conventional estimators:

1 The conventional metrics will provide biased estimates of distribution changes whenever sampling design has changed over time. A similar bias has been previously discussed in static habitat and climate envelope models (Kadmon, Farber & Danin 2004; Loiselle *et al.* 2008; Feeley & Silman 2010), where spatial variation in sampling intensity can lead to biased estimates of geographical distribution. However, spatial-varying sampling intensity has not to our knowledge been discussed as causing biased estimates of changes in species distribution over time. By contrast, our SDF estimator will generally be unbiased as long as changes in sampling intensity are statistically independent of changes in population abundance.

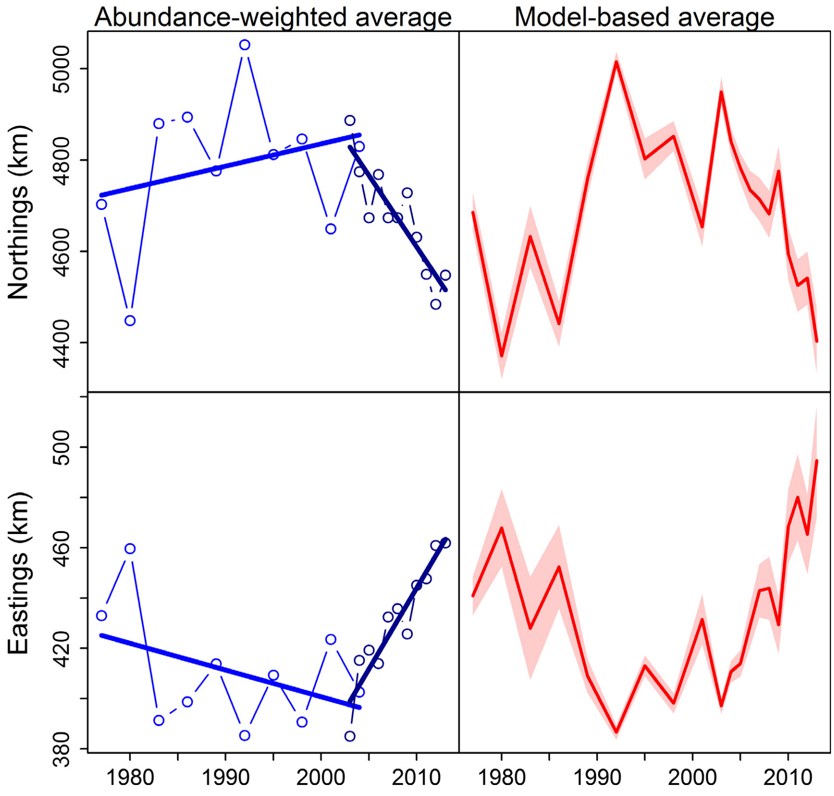

Fig. 4. Centre of gravity for Pacific hake shown in Universal Transverse Mercator projection using northings (top row) and eastings (bottom row), comparing abundance-weighted average (circles in left column, where light-blue indicates estimates from the triennial survey and dark-blue indicates estimates from the annual survey, and the corresponding blue line is the average trend from a linear regression model) and species distribution function approaches (where the thick red line is the maximum-likelihood estimate and the red shaded area is ±1 standard error).

2 The SDF estimator can incorporate information from multiple surveys simultaneously, as long as there is information to intercalibrate these surveys [e.g. samples from different surveys at nearby times and locations (Sauer, Peterjohn & Link 1994)].

3 The SDF approach can estimate standard error estimates for each year and therefore can be used to answer whether interannual variability exceeds the level that is expected given estimation error. By contrast, this task is difficult using

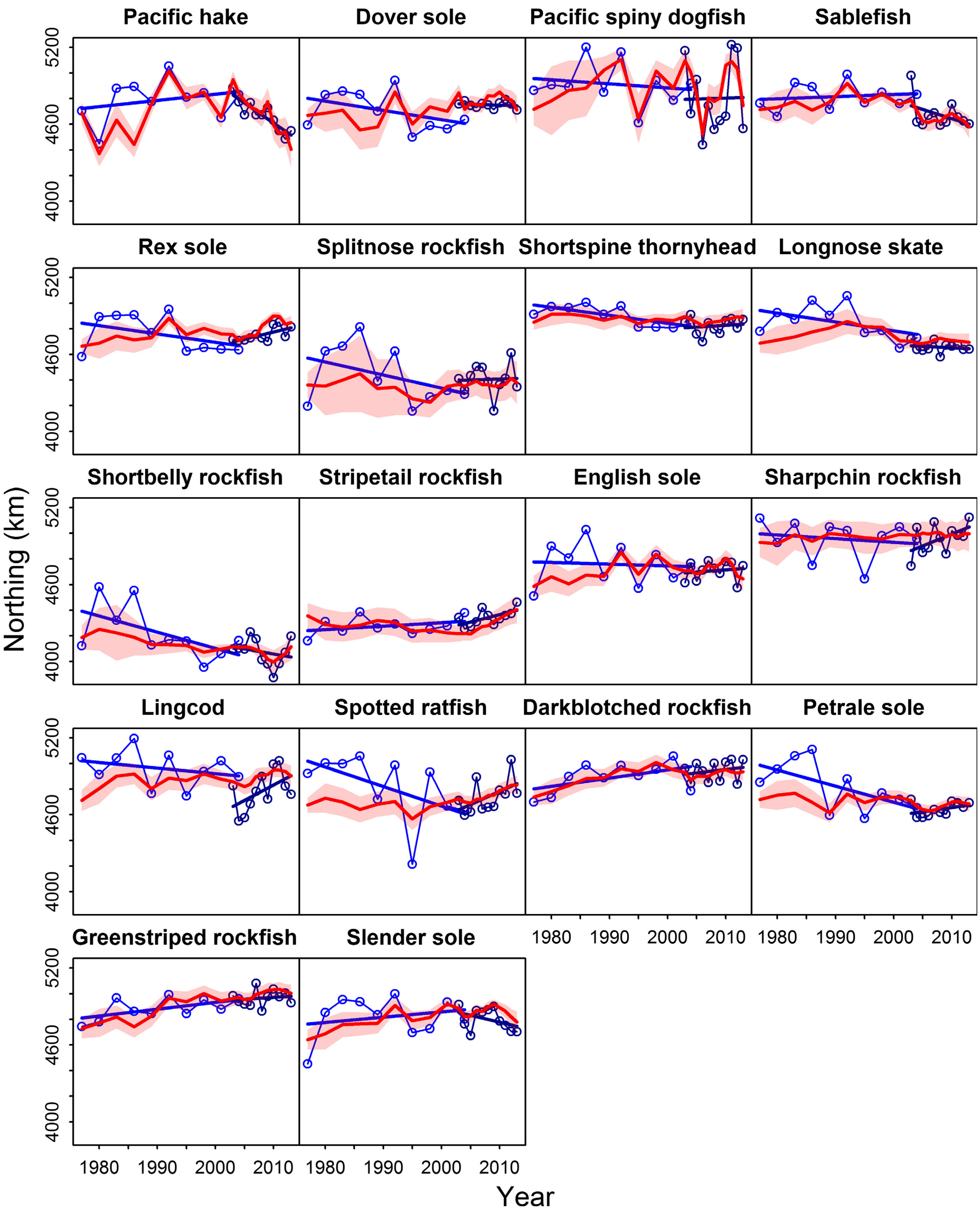

Fig. 5. A comparison abundance-weighted average and species distribution function estimators of northward distribution for all 18 West Coast species analysed in this study (see Fig. 3 caption for details).

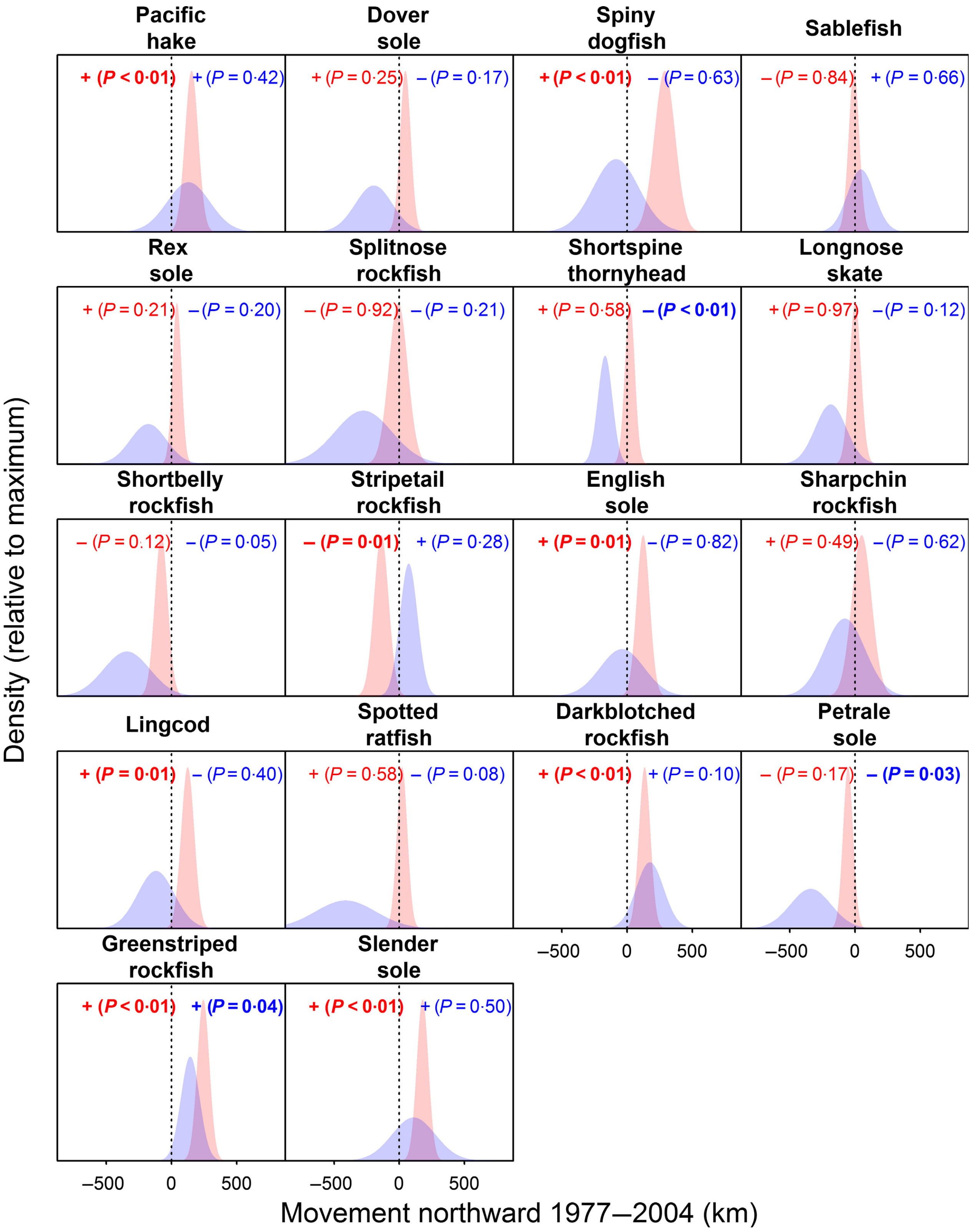

Fig. 6. Estimated change in northward centre of gravity using data from the triennial survey (1977–2004) and either the abundance-weighted average (blue) or new species distribution function (red) approaches, where we show a predictive distribution (centred on the maximum-likelihood estimates, with dispersion equal to the standard error) for each estimator, and where the sign (+: positive; –: negative) and *P*-value (from a Wald test) are shown for each estimator (red upper left: species distribution function; blue upper right: abundance-weighted average), and any *P*-value in bold font indicates a trend that is statistically significant using a two-sided Wald test at 0·05 significance.

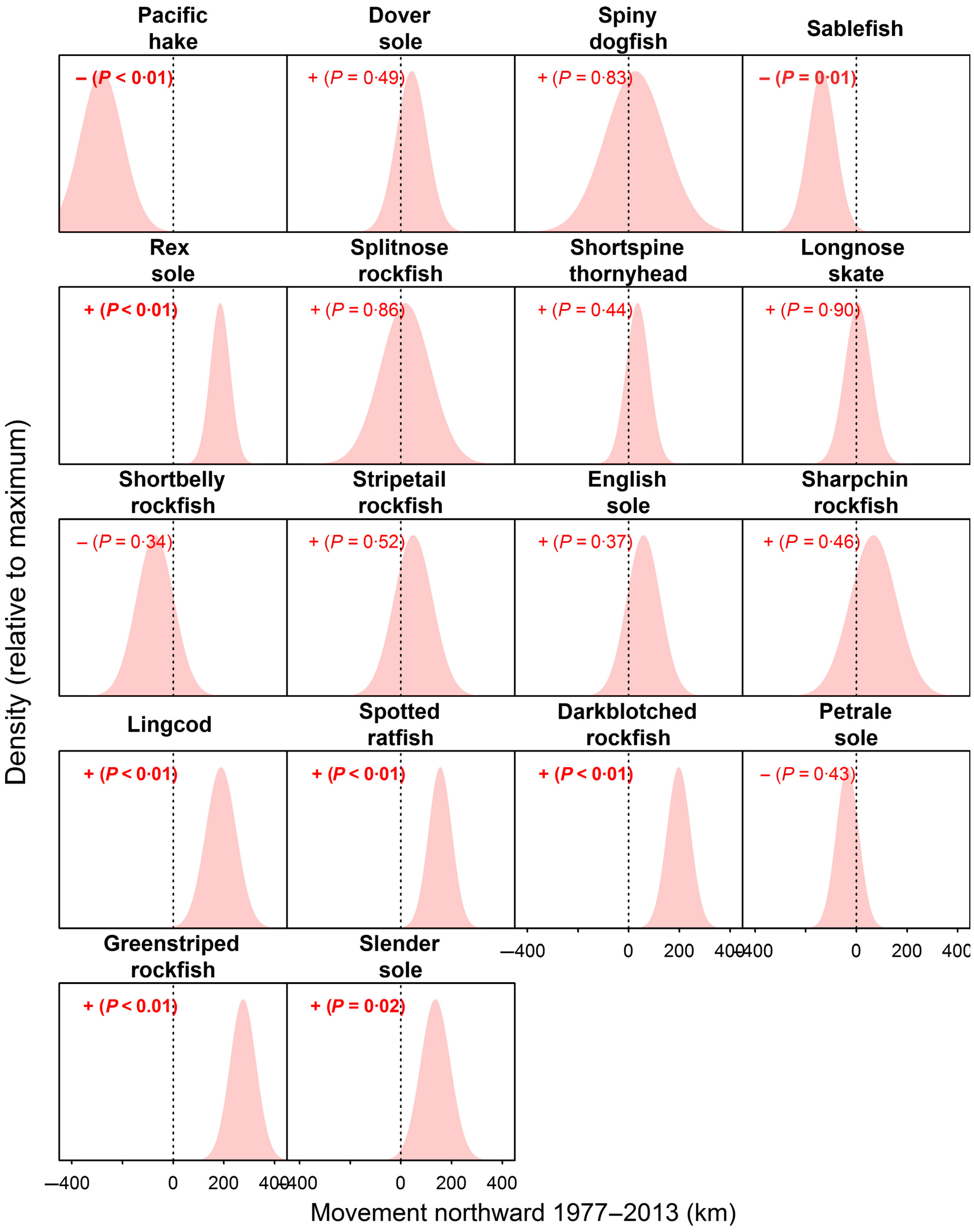

Fig. 7. Estimated change in northward centre of gravity using data from both triennial and annual surveys (1977–2013) and the new species distribution function (red) approaches, where we show a predictive distribution (centred on the maximum-likelihood estimates, with dispersion equal to the standard error) and where the sign (+: positive; −: negative) and *P*-value (from a Wald test) are shown in the upper left, and any *P*-value in bold font indicates a trend that is statistically significant using a two-sided Wald test at 0·05 significance.

conventional estimators, which are generally interpreted by fitting a linear model to estimate a single trend over time such that interannual residuals around this trend are often interpreted as statistical noise.

Estimated changes in species distributions can then be used to support hypotheses regarding causal environmental drivers, project future change (both in habitat utilization and the ecological niche that organisms inhabit) and

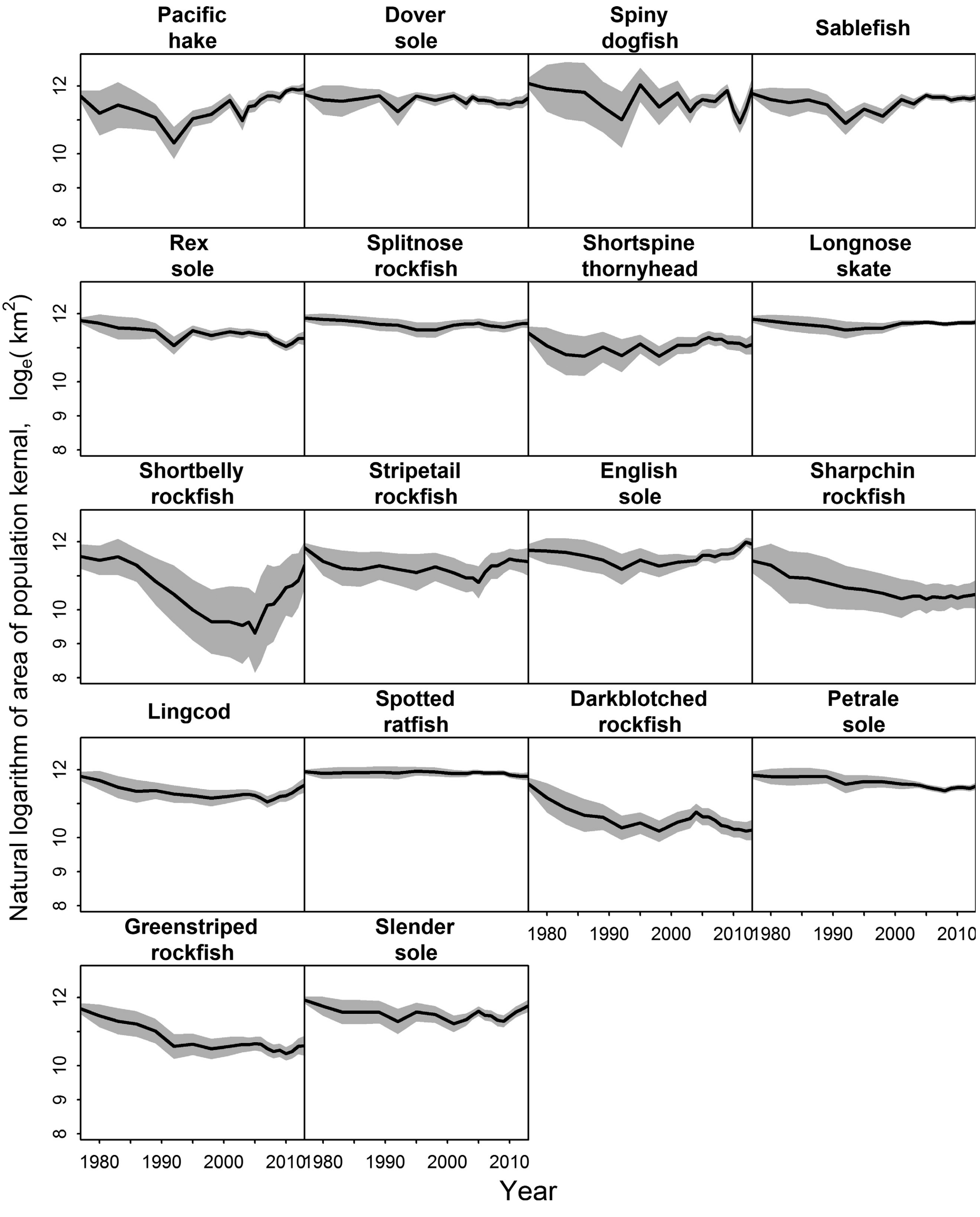

Fig. 8. Changes over time in the area occupied by each species, calculated as the volume the population kernel (an ellipse approximating the distribution that accounts for 50% of total population abundance, see Appendix S2) estimated from the population density function $d_p(s, t)$ for each species p and year t.

identify management actions that may mitigate against these changes.

Our case study application involving marine species on the west coast of the USA shows that a conventional estimator can lead to inference about range shifts in the opposite direction of inference from the SDF estimator. In particular, the southerly trend in the AWA estimator for some species reflects the southward shift in sampling that occurred during the triennial survey. The SDF estimator illustrates that pelagic species (Pacific hake and spiny dogfish) exhibit substantial variability in northward COG and that this variability exceeds the estimated standard error for each individual year. This result supports previous observations that Pacific hake distribution varies substantially over short time periods (Agostini *et al.* 2006). Spiny dogfish has also been documented previously to have demographics that vary over time (Taylor & Gallucci 2009), but spatial variation in distribution has received less attention than for Pacific hake. By contrast, rockfishes and thornyheads (*Sebastes* and *Sebastalobus* spp.) generally have very little interannual variability in north–south distribution. These species have widely varying life history, but generally have at least some association with demersal habitats that are thought to be relatively constant through time.

We have not linked our results explicitly to environmental variables, although marine species have previously been documented to shift in distribution with temperature changes (Pinsky *et al.* 2013; Hiddink, Burrows & García Molinos 2015). The SDF model that we use here could be linked to climate variables in at least three alternative ways. First, an environmental variable that varies spatially (e.g. local temperature at each location s and year t) could be included as a spatial covariate in the encounter probability or positive density functions (i.e. included in $x(s, t)$) if its value is known everywhere throughout the population domain (Shelton *et al.* 2014). This approach would permit the analyst to project changes in species distribution arising from different temperature scenarios that potentially vary spatially using output downscaled climate projection models. Secondly, an annual time series (e.g. average regional or global temperature) could be included as a covariate to predict changes in population-wide abundance, COG or width of the population kernel. This latter approach could be used to test, for example, whether changes in temperature lead to a more concentrated population distribution. Thirdly, temperature information could be used to compute local climate velocities (i.e. the ratio of temporal and spatial gradients in temperature at each location), and the spatial average of this local climate velocity could be compared with population-wide changes in COG to test whether the population is keeping pace with local temperature shifts (Devictor *et al.* 2008; Loarie *et al.* 2009). Each approach is suitable for different types of climate questions, and evaluating the relative merits of each approach remains an important topic for future research.

We recommend further research regarding distributional changes for rare or difficult-to-detect species. This can be accomplished by simultaneously estimating the distribution of rare and abundant species, such that estimates for rare species ‘borrow information’ from abundant species (Ovaskainen & Soininen 2011; Thorson *et al.* 2015a). We also recommend further research regarding age- and size-structured species distribution models. Many species including Pacific hake (Hicks *et al.* 2014), lingcod (Hamel, Sethi & Wadsworth 2009) and summer flounder (Bell *et al.* 2015) are known to have different spatial distribution for old and young individuals. Given the large variability in recruitment seen for most marine species (Thorson, Jensen & Zipkin 2014), fluctuations in recruitment can therefore drive large, interannual variation in the populations COG, where years following strong recruitment are biased towards the distribution for recruits. Fluctuations in distribution caused by variation in age structure might therefore be confused with climate-driven shifts in distribution, especially for short-lived species with strong cohorts (e.g. Pacific hake) or for long-lived species that rebuild their age structure following a reduction in fishing effort (e.g. summer flounder). Age- or size-structured distribution models could be estimated by either separately analysing the distribution for each age or size bin or jointly analysing them via some smoother across age or size (Kristensen *et al.* 2014). However, this may be complicated, given that individuals of different age or size often have different susceptibility to sampling gear. Nevertheless, we hypothesize that analysing climatic drivers of marine species distribution may be misleading unless care is taken to control for age-based distributional effects.

Acknowledgements

We thank Aurelie Godin, Anne Boudreau, Jim Hastie, Michelle McClure, Andrew Shelton and Jon Reum for helpful preliminary discussions, and the entire NWFSC assessment team for discussions about the shifts in sampling in the triennial survey. We also thank the many scientists and volunteers who contributed to the triennial and West Coast groundfish bottom trawl surveys. Finally, we thank Isaac Kaplan and two anonymous reviewers for helpful comments on an earlier draft.

Data accessibility

All data are publicly available from the Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA. The full and up-to-date data are available at https://www.nwfsc.noaa.gov/data/, and the data set used in this analysis (which involves a restricted spatial domain and set of species) is available from Dryad, http://dx.doi.org/10.5061/dryad.r1s8g.

References

Agostini, V.N., Francis, R.C., Hollowed, A.B., Pierce, S.D., Wilson, C. & Hendrix, A.N. (2006) The relationship between Pacific hake (*Merluccius productus*) distribution and poleward subsurface flow in the California Current System. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2648–2659.

Bahn, V. & McGill, B.J. (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, **16**, 733–742.

Bell, R.J., Richardson, D.E., Hare, J.A., Lynch, P.D. & Fratantoni, P.S. (2015) Disentangling the effects of climate, abundance, and size on the distribution of marine fish: an example based on four stocks from the Northeast US shelf. *ICES Journal of Marine Science: Journal du Conseil*, **72**, 1311–1322.

Bradburn, M.J., Keller, A.A. & Horness, B.H. (2011). *The 2003 to 2008 US West Coast bottom trawl surveys of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, length, and age composition.* US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA.

Cheung, W.W., Pinnegar, J., Merino, G., Jones, M.C. & Barange, M. (2012) Review of climate change impacts on marine fisheries in the UK and Ireland. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **22**, 368–388.

Cressie, N. & Wikle, C.K. (2011) *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Hoboken, NJ.

Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553–570.

Devictor, V., Julliard, R., Couvet, D. & Jiguet, F. (2008) Birds are tracking climate warming, but not fast enough. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**, 2743–2748.

Dulvy, N.K., Rogers, S.I., Jennings, S., Stelzenmüller, V., Dye, S.R. & Skjoldal, H.R. (2008) Climate change and deepening of the North Sea fish assemblage: a biotic indicator of warming seas. *Journal of Applied Ecology*, **45**, 1029–1039.

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

Engelhard, G.H., Righton, D.A. & Pinnegar, J.K. (2014) Climate change and fishing: a century of shifting distribution in North Sea cod. *Global Change Biology*, **20**, 2473–2483.

Feeley, K.J. & Silman, M.R. (2010) Land-use and climate change effects on population size and extinction risk of Andean plants. *Global Change Biology*, **16**, 3215–3222.

Fisher, J.A.D. & Frank, K.T. (2004) Abundance-distribution relationships and conservation of exploited marine fishes. *Marine Ecology Progress Series*, **279**, 201–213.

Garroway, C.J., Bowman, J., Cascaden, T.J., Holloway, G.L., Mahan, C.G., Malcolm, J.R., Steele, M.A., Turner, G. & Wilson, P.J. (2010) Climate change induced hybridization in flying squirrels. *Global Change Biology*, **16**, 113–121.

Hamel, O.S., Sethi, S.A. & Wadsworth, T.F. (2009). *Status and Future Prospects for Lingcod in Waters off Washington, Oregon, and California as Assessed in 2009*. *Pacific Fishery Management Council, Portland, Oregon, USA*.

Hicks, A.C., Taylor, N., Grandin, C., Taylor, I.G. & Cox, S. (2014). *Status of the Pacific Hake (Whiting) Stock in U.S. and Canadian Waters in 2013*. *International Joint Technical Committee for Pacific Hake, Seattle, Washington, USA*.

Hiddink, J.G., Burrows, M.T. & García Molinos, J. (2015) Temperature tracking by North Sea benthic invertebrates in response to climate change. *Global Change Biology*, **21**, 117–129.

Hitch, A.T. & Leberg, P.L. (2007) Breeding distributions of North American bird species moving north as a result of climate change. *Conservation Biology*, **21**, 534–539.

Kadmon, R., Farber, O. & Danin, A. (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401–413.

Kelly, A.E. & Goulden, M.L. (2008) Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11823–11826.

Kerr, J.T., Pindar, A., Galpern, P., Packer, L., Potts, S.G., Roberts, S.M. *et al.* (2015) Climate change impacts on bumblebees converge across continents. *Science*, **349**, 177–180.

Kristensen, K. (2014). *TMB: General Random Effect Model Builder Tool Inspired by ADMB*. https://cran.r-project.org/web/packages/TMB/index.html.

Kristensen, K., Thygesen, U.H., Andersen, K.H. & Beyer, J.E. (2014) Estimating spatio-temporal dynamics of size-structured populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 326–336.

Kristensen, K., Nielsen, A., Berg, C.W. & Skaug, H. (in press). Template model builder TMB. *Journal of Statistical Software*. http://arxiv.org/abs/1509.00660.

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

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The velocity of climate change. *Nature*, **462**, 1052–1055.

Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35**, 105–116.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J. & Possingham, H.P. (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**, 1235–1246.

Matthews, S.N., Iverson, L.R., Prasad, A.M. & Peters, M.P. (2011) Changes in potential habitat of 147 North American breeding bird species in response to redistribution of trees and climate following predicted climate change. *Ecography*, **34**, 933–945.

Ovaskainen, O. & Soininen, J. (2011) Making more out of sparse data: hierarchical modeling of species communities. *Ecology*, **92**, 289–295.

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

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J.K., Thomas, C.D., Descimon, H. *et al.* (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, **399**, 579–583.

Parmesan, C., Gaines, S., Gonzalez, L., Kaufman, D.M., Kingsolver, J., Townsend Peterson, A. & Sagarin, R. (2005) Empirical perspectives on species borders: from traditional biogeography to global change. *Oikos*, **108**, 58–75.

Perry, A.L., Low, P.J., Ellis, J.R. & Reynolds, J.D. (2005) Climate change and distribution shifts in marine fishes. *Science*, **308**, 1912–1915.

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

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rindorf, A. & Lewy, P. (2006) Warm, windy winters drive cod north and homing of spawners keeps them there. *Journal of Applied Ecology*, **43**, 445–453.

Sauer, J.R., Peterjohn, B.G. & Link, W.A. (1994) Observer differences in the North American Breeding Bird Survey. *The Auk*, **111**, 50–62.

Shelton, A.O., Thorson, J.T., Ward, E.J. & Feist, B.E. (2014) Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 1655–1666.

Taylor, I.G. & Gallucci, V.F. (2009) Unconfounding the effects of climate and density dependence using 60 years of data on spiny dogfish (*Squalus acanthias*). *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 351–366.

Thorson, J.T., Jensen, O.P. & Zipkin, E.F. (2014) How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 973–983.

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

Thorson, J.T., Shelton, A.O., Ward, E.J. & Skaug, H.J. (2015b) Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science: Journal du Conseil*, **72**, 1297–1310.

Thorson, J.T., Skaug, H.J., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J.H. & Benante, J.A. (2015c) The importance of spatial models for estimating the strength of density dependence. *Ecology*, **96**, 1202–1212.

Weinberg, K.L., Wilkins, M.E., Shaw, F.E. & Zimmermann, M. (2002). *The 2001 Pacific West Coast Bottom Trawl Survey of Groundfish Resources: Estimates of Distribution, Abundance, and Length and Age Composition*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, WA.

Woillez, M., Rivoirard, J. & Petitgas, P. (2009) Notes on survey-based spatial indicators for monitoring fish populations. *Aquatic Living Resources*, **22**, 155–164.

Received 30 November 2015; accepted 22 January 2016
Handling Editor: Olivier Gimenez

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Statistical properties of the abundance-weighted average estimator for center of gravity.

Appendix S2. Calculating center of gravity, population boundaries, the population kernel, and area occupied for the species distribution function model.

Appendix S3. Detailed description of methods for the spatiotemporal species distribution function model.

Appendix S4. Details regarding a simulation experiment evaluating the likely performance of AWA and SDF estimators.

Appendix S5. Visualizing the spatial distribution of sampling for triennial and annual surveys.

Appendix S6. Changes in the northern and southern population boundary for West Coast fishes.