

Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks

Brian C. Stock, Eric J. Ward, Tomoharu Eguchi, Jason E. Jannot, James T. Thorson, Blake E. Feist, and Brice X. Semmens

Abstract: Spatiotemporal predictions of bycatch (i.e., catch of nontargeted species) have shown promise as dynamic ocean management tools for reducing bycatch. However, which spatiotemporal model framework to use for generating these predictions is unclear. We evaluated a relatively new method, Gaussian Markov random fields (GMRFs), with two other frameworks, generalized additive models (GAMs) and random forests. We fit geostatistical delta-models to fisheries observer bycatch data for six species with a broad range of movement patterns (e.g., highly migratory sea turtles versus sedentary rockfish) and bycatch rates (percentage of observations with nonzero catch, 0.3%–96.2%). Random forests had better interpolation performance than the GMRF and GAM models for all six species, but random forests performance was more sensitive when predicting data at the edge of the fishery (i.e., spatial extrapolation). Using random forests to identify and remove the 5% highest bycatch risk fishing events reduced the bycatch-to-target species catch ratio by 34% on average. All models considerably reduced the bycatch-to-target ratio, demonstrating the clear potential of species distribution models to support spatial fishery management.

Résumé : Les prédictions spatiotemporelles de prises accessoires (c.-à-d. prises d'espèces non visées) sont prometteuses comme outils de gestion dynamique des océans pour réduire ces prises. Cependant, la nature du cadre de modélisation spatiotemporelle à utiliser pour produire ces prédictions n'est pas bien établie. Nous avons évalué une méthode relativement nouvelle, les champs aléatoires de Markov gaussiens (CAMG), et deux autres cadres, les modèles additifs généralisés (MAG) et les forêts aléatoires. Nous avons calé des modèles delta géostatistiques à des données d'observateurs des pêches sur les prises accessoires représentatives d'un vaste éventail de motifs de déplacement (p. ex. de tortues de mer très migratrices à des sébastes sédentaires) et de fréquences de prises accessoires (pourcentages d'observations de prise non nulle de 0,3 % à 96,2 %). Les forêts aléatoires donnent de meilleures interpolations que les modèles CAMG et MAG pour les six espèces, mais la performance des forêts aléatoires est plus sensible quand il s'agit de prédire des données à la limite de la pêche (c.-à-d. pour l'extrapolation spatiale). L'utilisation de forêts aléatoires pour cerner et retirer les 5 % des événements de pêche qui présentent le risque le plus élevé de prises accessoires réduit le rapport des prises accessoires et des prises de l'espèce visée de 34 % en moyenne. Tous les modèles réduisent considérablement ce rapport, ce qui démontre clairement le potentiel des modèles de distribution d'espèces comme outil pour appuyer la gestion spatiale des pêches. [Traduit par la Rédaction]

Introduction

Bycatch — catch of nontargeted species — occurs in nearly every commercial and recreational fishery and in many cases is a serious environmental and economic problem (Alverson et al. 1994; Davies et al. 2009; NMFS 2016). For high-profile protected species such as loggerhead sea turtles (*Caretta caretta*), even extremely low bycatch rates can result in population impacts and fisheries closures (Howell et al. 2015). Some species sustain highly valuable targeted fisheries but are considered bycatch in others, resulting in litigation and economic losses (e.g., Chinook salmon (*Oncorhynchus tshawytscha*) bycatch in the Bering Sea pollock (*Gadus chalcogrammus*) fishery; Ianelli and Stram 2015). Bycatch of undesired and unprotected species is also concerning because it reduces fishing efficiency and threatens ecosystem biodiversity (Boyce 1996; FAO 1995; Kelleher 2005). Thus, for a variety of rea-

sons, the fishing community is interested in tools to reduce bycatch.

One of these tools are maps of relative bycatch risk (e.g., probability or density) produced by species distribution models (SDMs). SDMs have seen rapid development in the last decade to meet critical conservation and resource management needs to understand how species distributions change in time and space (Parmesan and Yohe 2003; Sumaila et al. 2011; Pinsky et al. 2013). Accordingly, there is now a wide range of SDMs available to ecologists and fisheries scientists for fitting data on species presence-absence and abundance (Phillips et al. 2006; Illian et al. 2013; Conn et al. 2015; Golding and Purse 2016). SDMs have shown promise as tools for dynamic ocean management, which adapts to changing biological, oceanographic, or economic conditions faster than traditional, static, time, and area closures (Breivik et al. 2016; Dunn et al. 2016; Hazen et al. 2016; Howell et al. 2008, 2015). It is not

Received 18 July 2018. Accepted 5 June 2019.

B.C. Stock and B.X. Semmens. Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA.
E.J. Ward and B.E. Feist. Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanographic and Atmospheric Administration, 2725 Montlake Blvd. E, Seattle, WA 98112, USA.
T. Eguchi. Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.
J.E. Jannot and J.T. Thorson. Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 2725 Montlake Boulevard East, Seattle, WA 98112, USA.

Corresponding author: Brian C. Stock (email: b1stock@ucsd.edu).

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from [RightsLink](https://www.copyright.com).

clear, however, what SDM framework is most appropriate to use to support such tools. Further, because bycatch species vary from commonly to rarely caught, bycatch data sets offer a wide range of occurrence rates and densities. Thus, in addition to providing guidance for spatial bycatch management, large bycatch data sets are excellent testbeds for evaluating SDM performance more generally.

Species distribution models (SDMs)

SDMs can be coarsely divided into parametric, semiparametric, and nonparametric approaches. Generalized linear models (GLMs) are one of the simplest parametric approaches used to understand species distributions and their relationships with biotic and abiotic covariates (Venables and Dichmont 2004). GLMs predict the response variable, Y_i (species presence–absence or abundance at location i), by specifying a probability distribution and link function:

$$Y_i \sim \text{distribution with mean } \mu_i, \quad g(\mu_i) = \eta_i$$

with linear predictor

$$(1) \quad \eta_i = \mathbf{X}_i \boldsymbol{\beta}$$

where \mathbf{X}_i is a vector of covariate values for location i , and $\boldsymbol{\beta}$ is a vector of coefficients to be estimated. GLMs can permit nonlinear relationships between the covariates and response by including transformations of the covariates (e.g., polynomial terms $\eta_i = \beta_0 + \beta_1 \mathbf{X}_{i1} + \beta_2 \mathbf{X}_{i1}^2 + \beta_3 \mathbf{X}_{i1}^3 + \dots$) or by discretizing continuous covariates and treating them as categorical variables.

Generalized additive models (GAMs) extend the GLM framework by allowing the linear predictor to include smooth functions of the covariates (Guisan and Thuiller 2005; Wood 2017). GAMs are often referred to as semiparametric, since the smoothers do not have a specified functional form but do have associated parameters that are estimated using penalized likelihood (Wood 2011; Guélat and Kéry 2018). The ability of GAMs to incorporate complex, nonlinear covariate effects, as well as improvements to computing power and software, has led to their wide adoption in fisheries and ecology in the last decade (Becker et al. 2014; Leathwick et al. 2006; Li and Pan 2011; Watson et al. 2009). Extending the linear predictor in eq. 1 to include a two-dimensional spline, f , on the geographical coordinates of location i , \mathbf{s}_i , specifies a GAM:

$$(2) \quad \eta_i = \mathbf{X}_i \boldsymbol{\beta} + f(\mathbf{s}_i)$$

Equation 2 is estimated by penalized likelihood maximization, which balances smoothness and fit to the data by penalizing the curvature (i.e., integral of the squared second derivative) of $f(\mathbf{s}_i)$ (Wood 2017). Kammann and Wand (2003) refer to eq. 2 as a “ge-additive” model and have shown that this is mathematically equivalent to explicitly modeling spatial correlation with random effects, \mathbf{u} (Diggle et al. 1998; Kneib et al. 2008; Péron et al. 2011; Fahrmeir et al. 2013; Guélat and Kéry 2018):

$$(3) \quad \eta_i = (\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i)$$

When the spatial random effects are assumed to follow a zero-mean multivariate normal distribution (MVN), eq. 3 can be written as

$$(4) \quad \eta_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}(\mathbf{s}_i), \quad \boldsymbol{\varepsilon}(\mathbf{s}) \sim \text{MVN}(0, \boldsymbol{\Sigma})$$

where $\boldsymbol{\varepsilon}$ is a Gaussian field (Kneib et al. 2008). Analogous to how the curvature of the spline is penalized when estimating the GAM, the correlation function that defines $\boldsymbol{\Sigma}$ acts as a penalized spatial smoother in the Gaussian field model — nearby locations are more highly correlated, and thus more smoothed, than distant locations (Fahrmeir et al. 2013). Gaussian fields are attractive because they directly model spatial correlation, but applications have historically been limited to smaller data sets because inverting the covariance matrix $\boldsymbol{\Sigma}$ makes them computationally intense (Lindgren et al. 2011). In summary, both the GAM and Gaussian field model account for spatial autocorrelation not explained by environmental covariates, are semiparametric, mathematically equivalent, and typically fit using penalized likelihood that optimizes spatial smoothing.

Gaussian fields are defined in continuous space but can be approximated by discrete Gaussian Markov random fields (GMRFs; Lindgren et al. 2011). GMRFs have increasingly been used to model species distributions as advances in computing power and software implementation have allowed ecologists to apply them to large data sets. Among other advantages, the GMRF approach can be implemented by integrated nested Laplace approximation, which is faster than other methods of Bayesian inference (i.e., Markov chain Monte Carlo) and allow GMRF approximations of Gaussian fields to be computationally feasible (Rue et al. 2009). GMRF models have shown promise in assessing relationships between habitat and distribution (Illian et al. 2013), the effects of interspecific relationships such as density dependence (Thorson et al. 2015c), as well as the relationships between multiple co-occurring taxa (Ward et al. 2015). From a quantitative standpoint, GMRF models have been shown to estimate population abundance trends with greater precision and accuracy compared with nonspatial models (Thorson et al. 2015b).

Clearly, GMRFs are intimately related to GAMs, since GMRFs approximate the Gaussian field model (eq. 4), which is an alternative parameterization of the GAM in eq. 2. The spatial smoothing terms for both GAMs and GMRFs can be defined on a sphere or differ according to spatial direction (anisotropy can be included in GAMs by using tensor product smooths, and $\boldsymbol{\Sigma}$ can be both nonstationary and anisotropic in GMRFs). GAMs and GMRFs differ, however, in two regards. First, the spatial smoothing term appears as a mean trend for GAMs and as a covariance matrix for GMRFs. This distinction may not be important for modeling species distributions, and researchers may prefer the method that reflects their view of how spatial autocorrelation arises in the problem at hand. For instance, GMRFs would be the more natural framework if the spatial variation remaining after including environmental covariates is considered random. Second, and more importantly, the different parameterizations lead to different estimation methods and software implementations. Since spatial models for large data sets contain many parameters, numerically efficient implementations are crucial (Fahrmeir et al. 2013). Several R packages fit GAMs, the most popular being “mgcv”, which uses generalized cross-validation to estimate smooth terms by default (Wood 2017). GMRFs can be fit via “INLA”, which uses a Bayesian framework and estimates models by integrated nested Laplace approximation (Lindgren and Rue 2015). In theory, “mgcv” and “INLA” should be quite similar. In practice, however, differences in approximation methods, runtime, convergence criteria, ease of use, and default settings may impact model predictions.

Most nonparametric approaches ecologists use to model species distributions have evolved from machine learning algorithms (Hastie et al. 2009; Olden et al. 2008). These data-driven approaches include random forests (RF; Breiman 2001; Cutler et al. 2007), MaxEnt (Phillips et al. 2006; Phillips and Dudík 2008), and support vector machines (Drake et al. 2006). In this analysis, we highlight RF because (i) data in our application — fisheries bycatch — contain true absences, whereas MaxEnt is designed

for presence-only data, and (ii) RF is widely used and has shown good predictive performance in SDM testing (Prasad et al. 2006; Marmion et al. 2009; Scales et al. 2016). The RF algorithm predicts the response by constructing m regression (or classification) trees and averaging their predictions (Breiman 2001). Each individual tree begins with all observations and then iteratively partitions the data by splitting along one covariate (e.g., depth > 100 m versus depth \leq 100 m), choosing the covariate and split point that minimizes the sums of squares error at each node (where the predicted response at each node is the mean of observations within the node; Breiman et al. 1984). The process continues until each terminal node contains less than a specified number of observations. Individual trees are simple and computationally cheap but are also unstable (i.e., sensitive to slight alterations in the data) and suboptimal at prediction (i.e., they are “weak learners”), because they only allow rectangular partitions of covariate space. The RF algorithm increases predictive performance by reducing the correlation between trees, which is accomplished via two processes: (i) fitting each tree to a bootstrap sample of the original data and (ii) at each split, randomly selecting a subset of covariates to consider (Kuhn and Johnson 2013). This works because reducing the correlation between individual trees reduces the correlation of their errors, which therefore reduces the predictive error of their average, the RF estimate.

Random forests are popular because they are simple to use (few parameters to tune and the default values work well in most cases), robust to the inclusion of many noninformative covariates generate accurate predictions, designed to not overfit, and seamlessly accommodate missing data (Biau and Scornet 2016). Compared with parametric and semiparametric models, RF will often have better out-of-sample (i.e., cross-validated) prediction performance due to their ability to estimate more complex patterns, as nonlinearity and interactions are inherent in their construction (Elith and Leathwick 2009). However, this data-driven complexity does come at the cost of model interpretability, and this is one of the main factors limiting the adoption of RF — and machine learning methods more generally — by ecologists (Olden et al. 2008). Three other disadvantages of RF are the difficulty of generating uncertainty estimates with well-understood properties, analyzing model diagnostics, and specifying constraints on model fit (e.g., we may wish yearly estimates to be independent, which can be specified in parametric models).

Study objectives

The primary objective of this paper is to compare the performance of GAMs, GMRFs, and RF in a predictive framework using cross-validation (Kuhn and Johnson 2013; Roberts et al. 2017). There has been an increased emphasis in ecology on evaluating and selecting models based on their ability to predict out-of-sample data (Hooten and Hobbs 2015), and one of the advantages of this approach is that nonparametric and parametric models can be compared (Ward et al. 2014). While each of these model frameworks have individually been applied to understand spatiotemporal trends in fisheries bycatch (GAMs: Becker et al. 2014; Hazen et al. 2016; McCracken 2004; Watson et al. 2009; GMRFs: Breivik et al. 2016, 2017; Cosandey-Godin et al. 2015; RFs: Carretta et al. 2017; Eguchi et al. 2017; Pons et al. 2009), their predictive performance has not been tested in a comparative study.

Our next objective is to evaluate the utility of using SDM predictions of bycatch risk as a tool to reduce bycatch in fisheries. Beyond abstract performance metrics, we compare the models' capabilities to reduce the bycatch-to-target species catch ratio, create spatial bycatch risk maps, and estimate effects of covariates.

The final objective of our analysis is to evaluate model transferability, the ability to extrapolate or predict beyond the range of observed data. Traditional cross-validation only measures a model's ability to interpolate (i.e., estimate values within the

range of observations), because it randomly chooses data to withhold for testing. SDMs that are more data-driven and complex have been shown to have better interpolation performance but be worse at spatial extrapolation (Araújo and Rahbek 2006; Heikkinen et al. 2012; Randin et al. 2006). In other words, one model may have higher predictive performance in the core fishing area with abundant data, yet underperform other models in areas with sparse sampling coverage. Since we wish to evaluate using SDM predictions of bycatch risk as a spatial management tool, it is important to assess how sensitive the predictions are to spatial location. Predictions in areas with few data are more sensitive to model misspecification and overfitting, and therefore caution is especially warranted for complex, nonparametric approaches such as RF (Merow et al. 2014).

Methods

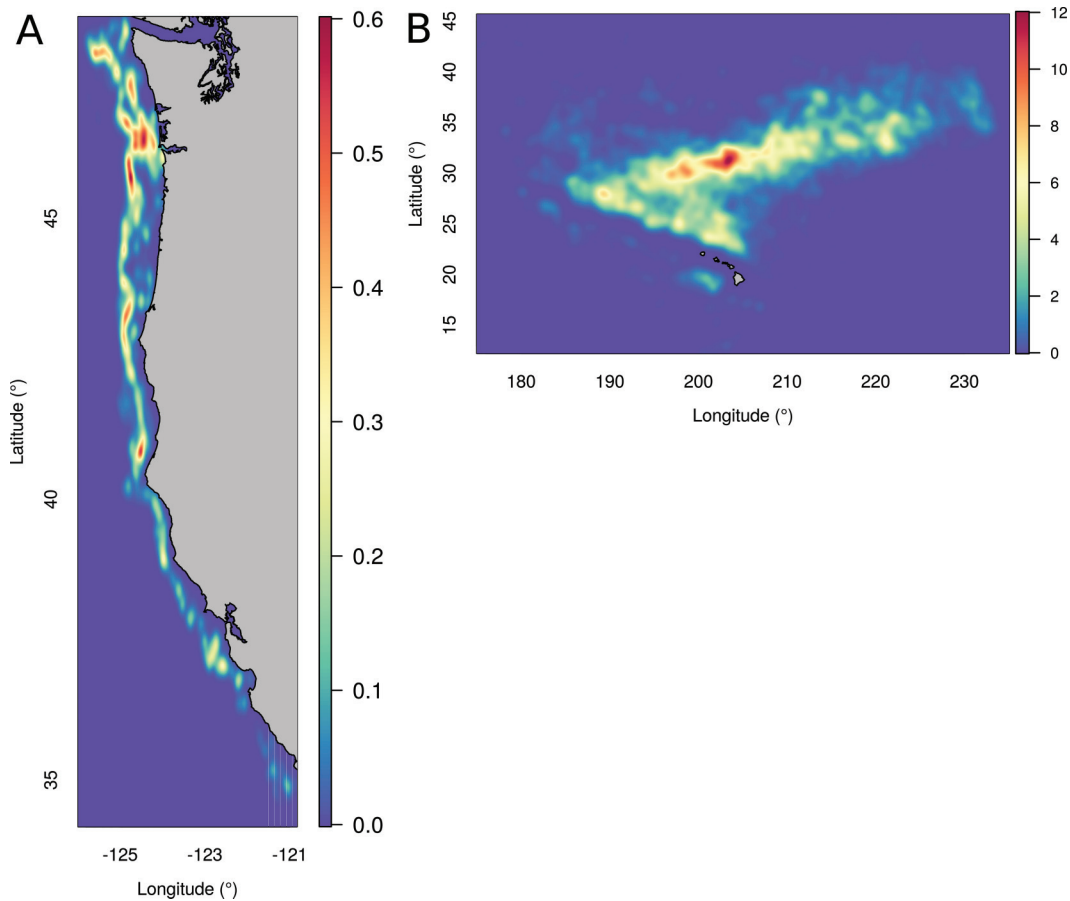
Fisheries observer data

Collecting reliable bycatch data depends on fisheries observer programs, where onboard observers enumerate and record the species caught (as well as fishing location, gear type, time, and other relevant information). To explore the performance of species distribution models across taxa, we used two data sets from United States fisheries observer programs in the Pacific Ocean with high observer coverage. The first data set was from the West Coast Groundfish Observer Program (WCGOP) at the Northwest Fisheries Science Center (NWFSC) (Bellman et al. 2010). The WCGOP data set contained records of 42 786 commercial bottom trawls from 2003 to 2012 off the West Coast of the USA, primarily targeting groundfish such as Dover sole (*Microstomus pacificus*), thornyheads (*Sebastolobus* spp.), sablefish (*Anoplopoma fimbria*), and rockfish (*Sebastes* spp.; Fig. 1A). Observers recorded haul duration, location, date, time, depth, gear type, and catch (which includes at-sea discarded bycatch; for details see NWFSC 2016). Observer coverage was ~20% from 2003 to 2010 under limited access management, with 100% coverage starting in 2011 with the transition to an individual fishing quota (IFQ) system. In the pre-IFQ era, fishermen were not permitted to land rebuilding species (i.e., populations declared overfished with management plans to rebuild to sustainable levels), so we defined bycatch as only at-sea discards. Under the IFQ system, fishermen can land a low quota of rebuilding species, so we considered bycatch to be the sum of discarded and retained catch for nontarget species.

The second data set was from the Hawaii longline (HILL) fishery, monitored by the Pacific Islands Regional Observer Program (PIROP 2014), which has recorded fishing location, date, time, sea surface temperature (SST), gear characteristics, and catch of longline sets from 1994 to 2014. The Hawaii longline fleet is divided into two sectors, one targeting tuna (*Thunnus* spp.) and the other swordfish (*Xiphias gladius*), with distinct gear configurations and spatiotemporal effort patterns, both of which affect interaction rates with bycatch species (Li and Pan 2011). We modeled 16 714 observations from the shallow-set swordfish fishery in 1994–2001 and 2005–2014 (Fig. 1B), distinguishing between sets targeting swordfish and tuna by the number of hooks between surface floats (following Li and Pan 2011). Concerns over bycatch of protected species, particularly of loggerhead (*Caretta caretta*) and leatherback (*Dermochelys coriacea*) sea turtles, motivated the closure of the swordfish fishery from 2001 to 2004. This led to two important differences between the data for 1994–2001 and 2005–2014. First, sea turtle bycatch rates have been an order of magnitude lower in the later period, the result of stricter regulations and modifying hooks (J to circle hooks) and bait types (squid to fish; Gilman et al. 2007). Second, observer coverage increased from roughly 5% to 100% (Howell et al. 2008).

Model performance may be linked to species' movement patterns, because species that move less (or whose movement patterns do not change in time) may not need a spatiotemporal

Fig. 1. Spatial extent of the two fisheries observer data sets. (A) Fishing effort in the West Coast groundfish trawl fishery from 2003 to 2012 (42 786 haul locations). (B) Fishing effort in the shallow-set Hawaii longline swordfish fishery from 1994 to 2014 (16 714 set locations). Bivariate kernel density estimates of fishing effort were used to smooth the data (“bkde2D” function in R package “KernSmooth”). [Colour online.]



model. Instead, a time-constant spatial model may be adequate. To ascertain whether differences in SDM performance were related to movement pattern or bycatch rate (i.e., percent observations with nonzero catch), we selected three bycatch species from each data set: blue shark (*Prionace glauca*), loggerhead sea turtle, and leatherback sea turtle from the Hawaii longline fishery, and Pacific halibut (*Hippoglossus stenolepis*), darkblotched rockfish (*Sebastes crameri*), and yelloweye rockfish (*Sebastes ruberrimus*) from the West Coast groundfish trawl fishery. These species widely differ in their bycatch rates (96.2%, 0.7%, 0.3%, 28.9%, 17.9%, and 1.4%, respectively), habitat preferences, and movement patterns. For instance, rockfish are relatively sedentary and closely associated with rocky bottom habitat, whereas halibut exhibit seasonal and long-distance migrations (Skud 1977; Gunderson 1997). In contrast with the groundfish, blue sharks and sea turtles inhabit the open ocean and range much more widely (Benson et al. 2011; Kobayashi et al. 2008; Nichols et al. 2000).

While both data sets include periods with 100% observer coverage, they also span periods with partial coverage. This is relevant since the models assume that the data represent a random sample of the studied fishery (i.e., each fishing event has an equal probability of being observed). For several reasons, it is difficult for observer programs to achieve random sampling: a list of trips and their departures often does not exist far in advance, certain vessels may not be able to accommodate observers, observers may not always be available, and fisher behavior can change when observers are on board (Hall 1999; Liggins et al. 1997; McCracken 2004). The WCGOP data from years with 20% coverage are likely to be representative of the fishery, because the WCGOP stratified

sampling by port group, vessel, and 2-month blocks with the goals of sampling all vessels for 2 months in each year and discouraging changes to fishing behavior when observers were onboard (NWFSC 2006). It is less likely that this was true for the 1994–2001 HILL data. Nevertheless, we included data from periods with partial coverage because there were very few observations of nonzero catch for rarely encountered species in the years with full coverage (yelloweye rockfish: 38, loggerhead turtle: 89, leatherback turtle: 82), and in many cases, bycatch of these “rare-event” species are often of highest management concern (Martin et al. 2015).

Environmental covariates

In addition to the locations of observed fishing, we considered several covariates that may help explain the likelihood of bycatch events. For the WCGOP data set, we included fishing depth, day of year, SST anomaly, distance to rocky habitat, size of nearest rocky patch, predicted occurrence from survey data, and whether the trawl occurred in or near a Rockfish Conservation Area (RCA). RCAs are large areas along the US West Coast closed to fishing designed primarily to reduce bycatch of overfished rockfish, such as two of the species we considered. RCA boundaries have changed by and within years and are defined by latitude, date, and depth (NOAA Fisheries West Coast Region 2015). Trawls were determined to be inside or outside of an RCA based on the trawl date, average start and end trawl position, and bottom depth (calculated via bathymetry from NOAA National Centers for Environmental Information 2015). We included linear and quadratic terms for fishing depth and SST anomalies following Shelton et al. (2014). Depth was recorded by onboard observers,

Table 1. Properties of the considered statistical models and how each model incorporates spatial fishing locations.

Model	Parametric?	Computational intensity	R package	Inclusion of spatial locations
Generalized linear model (GLM)	Parametric			
GLM		Low	mgcv	None
Generalized additive model (GAM)	Semiparametric			
GAM-CONSTANT		Low	mgcv	+s(Lat, Lon, k = 100)
GAM-YEAR		Medium	mgcv	+s(Lat, Lon, k = 100, by = year)
Gaussian Markov random field (GMRF)	Semiparametric			
GMRF-CONSTANT		High	INLA	+f(i, model = spde)
GMRF-YEAR		Very high	INLA	+f(i, model = spde, group = year, control.group = list(model = "exchangeable"))
Random forest (RF)	Nonparametric			
RF-BASE		Low	randomForest	+Lat+Lon
RF-DOWN		Low	randomForest	+Lat+Lon
RF-SMOTE		Low	caret	+Lat+Lon

Note: The GLM model serves as the baseline model — no spatial data included. GAM models fit two-dimensional splines on geographical coordinates (i.e., latitude and longitude), either constant across years (GAM-CONSTANT) or estimating a different spline for each year (GAM-YEAR). GMRF models incorporate spatial locations by estimating the covariance between locations as a random field (with stationary Matern covariance function). As for GAMs, we fit GMRFs that estimate one random field kept constant across years (GMRF-CONSTANT) or estimate a random field for each year (GMRF-YEAR). RF is nonparametric and thus only incorporates spatial locations by including covariates of latitude and longitude. All models for a given species were fit using the same nonspatial covariates (habitat, depth, SST, etc.). We used the R packages “mgcv” (GLM and GAM), “INLA” (GMRF), “randomForest” (RF), and “caret” (RF-SMOTE).

while SST anomalies were measured via satellite. For each trawl, we collected daily SST anomalies on a 0.25° grid and used bilinear interpolation to create SST anomalies corresponding to each trawl location (<http://www.esrl.noaa.gov/psd/>; Reynolds et al. 2007). Rocky habitat data were from NMFS (2013), calculated as per Shelton et al. (2014). Finally, we used the above covariates to fit a geostatistical binomial GLMM to fisheries-independent trawl survey data (Bradburn et al. 2011, modeled as in Shelton et al. 2014) and applied this model to predict bycatch occurrence at the fishing times and locations in the observer data set. These survey-predicted occurrence probabilities were included as another linear covariate. All environmental covariates were centered before model estimation.

The only available environmental covariate for the HILL data set was observer-recorded SST, and therefore we fit the HILL models with covariates of standardized SST, SST², and day of year.

Statistical models

As is common for species distribution data, five of the six species exhibited large proportions of zero catches. We followed the hurdle or delta-model approach to this complication, which is commonly applied in ecology and fisheries (Pennington 1983; Maunder and Punt 2004). Delta-models separate the observed catches, Y_i , into two processes: a “binomial” component for the probability of nonzero catch, π_i , and a “positive” component for the mean catch density given the catch is nonzero, μ_i :

$$(5) \quad \begin{aligned} Z_i &\sim \text{Bernoulli}(\pi_i) \\ Y_i &\sim Z_i h(\mu_i) \end{aligned}$$

where Z_i is a binary variable that equals 1 if the species was caught and 0 if it was not, and $h()$ is a distribution to be specified (e.g., lognormal, gamma). Splitting the modeling into these two components can be advantageous because different mechanisms may affect one component but not the other (e.g., a habitat quality covariate may be a significant predictor of catch rate, but not occurrence).

We applied a total of eight delta-models with varying spatial structure to each of the six species included in our analysis (Table 1). Bycatch of yelloweye rockfish, loggerhead turtles, and leatherback turtles were extremely rare events (0.3%–1.4% nonzero observations), with too few multiple-individual catches to meaningfully fit the positive component. All analyses were conducted using R version 3.4.1 (R Core Team 2017), with the follow-

ing libraries: “mgcv” was used to implement GLMs and GAMs (version 1.8-17; Wood 2017); “randomForest” (version 4.6-12; Liaw and Wiener 2002), “DMwR” (version 0.4.1; Torgo 2010), and “forestFloor” (version 1.9.5; Welling et al. 2016) were used to fit RFs; and “INLA” was used to fit the GMRF models (version 0.0-1485844051; Lindgren and Rue 2015). We assessed model fit with plots of covariate-response relationships, predicted versus observed response in out-of-sample data, spatial residual maps, and spatial correlograms (Moran’s I , package “nfc” version 1.2-5; Bjørnstad and Falck 2001). Code to fit each of the models is provided at <https://github.com/brianstock/spatial-bycatch>.

Our first model was a delta-GLM with linear and quadratic effects of the environmental covariates (which are intrinsically spatially correlated), but without any spatial terms — neither geographic coordinates nor spatial autocorrelation for residual errors. As in Guélat and Kéry (2018), the delta-GLM served as a baseline that allowed us to evaluate the value of adding spatial terms in the subsequent models, which were fit using the same covariates and only differ in how they include spatial information. The delta-GLM fits the observed bycatch in fishing event i , Y_i , as in eqs. 1 and 5, with the binomial component determining the probability of nonzero bycatch, π_i :

$$(6) \quad Z_i \sim \text{Bernoulli}(\pi_i), \quad \text{logit}(\pi_i) = \mathbf{X}_i \boldsymbol{\alpha}$$

and positive component for the mean catch density given the catch is nonzero, μ_i :

$$(7) \quad Y_i \sim Z_i \text{Gamma}(\mu_i, k), \quad \log(\mu_i) = \mathbf{X}_i \boldsymbol{\beta}$$

where \mathbf{X}_i is a vector of covariate values for location i , $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of coefficients to be estimated, and k is the shape parameter of the gamma distribution. The gamma distribution is appropriate for positive, right-skewed data and therefore is commonly used in the positive component of delta-models for fisheries catch (Lecomte et al. 2013; Stefánsson 1996). While we would not expect the GLM to outperform the models with explicit spatial terms, it is possible that the (spatially structured) environmental covariates could explain most of spatial structure in the response. In that case, including spatial terms in the model (i.e., a two-dimensional spline as in eq. 2 or covariance matrix as in eq. 4) would be unnecessary.

We fit two delta-GAM models that extend eqs. 6 and 7 by adding a two-dimensional spline, $f()$, on the geographical coordinates of location i , \mathbf{s}_i , to both the binomial and positive components, as in eq. 2:

$$(8) \quad \begin{aligned} Z_i &\sim \text{Bernoulli}(\pi_i), & \text{logit}(\pi_i) &= \mathbf{X}_i\alpha + f_Z(\mathbf{s}_i) \\ Y_i &\sim \text{Gamma}(\mu_i, k), & \text{log}(\mu_i) &= \mathbf{X}_i\beta + f_Y(\mathbf{s}_i) \end{aligned}$$

The first, “GAM-CONSTANT”, includes one two-dimensional spline constant across years, with an offset (fixed effect) for each year. This allows the mean bycatch probability and density to vary temporally and spatially, but in the same pattern each year. The second, “GAM-YEAR”, fits an independent two-dimensional spline for each year, which allows the spatial pattern to vary between years (Table 1).

As for the GAMs, we fit two delta-GMRF models that extend eqs. 6 and 7 by estimating the covariance between observed locations, \mathbf{s}_i , as in eq. 4:

$$(9) \quad \begin{aligned} Z_i &\sim \text{Bernoulli}(\pi_i), & \text{logit}(\pi_i) &= \mathbf{X}_i\alpha + \varepsilon_Z(\mathbf{s}_i), & \varepsilon_Z(\mathbf{s}) &\sim \text{MVN}(0, \mathbf{Q}_Z^{-1}) \\ Y_i &\sim \text{Gamma}(\mu_i, k), & \text{log}(\mu_i) &= \mathbf{X}_i\beta + \varepsilon_Y(\mathbf{s}_i), & \varepsilon_Y(\mathbf{s}) &\sim \text{MVN}(0, \mathbf{Q}_Y^{-1}) \end{aligned}$$

where both \mathbf{Q}_Z^{-1} and \mathbf{Q}_Y^{-1} are defined to approximate stationary, isotropic Matérn covariances:

$$\text{Cov}(\mathbf{s}_1, \mathbf{s}_2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa\|\mathbf{s}_1 - \mathbf{s}_2\|)^{\nu} K_{\nu}(\kappa\|\mathbf{s}_1 - \mathbf{s}_2\|)$$

K_{ν} is the modified Bessel function of the second kind and order $\nu > 0$, κ is the spatial scale parameter, and $\varepsilon_Z()$ and $\varepsilon_Y()$ represent the estimated spatial fields using random effects. We used the default Matérn smoothness, $\nu = 1$, and priors on parameters as implemented in R-INLA (Lindgren and Rue 2015). Analogous to the GAM-CONSTANT and GAM-YEAR models, we fit a “GMRF-CONSTANT” model with one random field constant across all years and a “GMRF-YEAR” model with a random field estimated for each year (Table 1). As for GAM-CONSTANT, the GMRF-CONSTANT model includes fixed effect terms for each year, which allow for an increase or decrease in the mean bycatch probability and density for each year while assuming the spatial pattern is constant across years. The GMRF-YEAR model uses the simplest spatiotemporal option in R-INLA, “exchangeable”, which refers to the spatiotemporal structure — the random fields in all years are uniformly correlated (as opposed to an autoregressive spatiotemporal structure where nearby years are more correlated than distant years).

To include spatiotemporal effects in RFs, we added year (treated as a factor), latitude, and longitude as covariates. For the positive component of the delta-model, we fit only one RF model: “RF-BASE,” following the original RF algorithm as described by Breiman (2001) and implemented in the “randomForest” R package (Liaw and Wiener 2002). For the binomial component, we also fit two modifications to the original RF algorithm designed to improve performance on imbalanced class data (i.e., proportions of 0s and 1s very unequal), because several species showed strong class imbalance (e.g., yelloweye rockfish had 99.7% tows with zero catch and only 0.3% tows with nonzero catch). Training an RF on such severely imbalanced class data tends to produce models that predict the majority class well but performs poorly on the minority class (Kuhn and Johnson 2013). The first approach was to down-sample the majority class observations (e.g., 0s, tows with zero catch for yelloweye rockfish) when training the RF such that each tree used a stratified bootstrap sample of the data with equal numbers of majority and minority (e.g., 1s, tows with nonzero catch for yelloweye rockfish) class observations. We implemented

down-sampling in the “RF-DOWN” model, setting the sample size equal to one-sixth the number of minority class observations. The second approach, “RF-SMOTE” (synthetic minority over-sampling technique; Chawla et al. 2002), combined down-sampling of the majority class with over-sampling of the minority class. Simply over-sampling the minority class with replacement overfits the RF to the specific observed minority data and typically does not significantly improve prediction of the minority class. Instead, SMOTE creates synthetic minority class samples by generating random linear combinations of nearby observed minority samples (i.e., if X_1 and X_2 are nearest neighbors, with $X_1 = (\text{SST}_1, \text{Lat}_1, \text{Lon}_1)$ and $X_2 = (\text{SST}_2, \text{Lat}_2, \text{Lon}_2)$, draw p from Uniform(0, 1) and create $X_{\text{syn}} = [p(\text{SST}_1 - \text{SST}_2) + \text{SST}_2, p(\text{Lat}_1 - \text{Lat}_2) + \text{Lat}_2, p(\text{Lon}_1 - \text{Lon}_2) + \text{Lon}_2]$).

Performance metrics

We performed fivefold cross-validation repeated 10 times, which allowed us to evaluate semiparametric and nonparametric SDMs’ predictive error on new data (Shmueli 2010; Kuhn and Johnson 2013). We blocked by year (rather than systematically excluding a given year) to account for temporal structure and mimic predictive performance for cases where a random subset of samples in a given year are not observed (Roberts et al. 2017). Thus, we generated 50 test–train splits, where each selects 20% of the data from each year to reserve for testing and fits the models on the remaining 80% training data. After fitting the models to training data, we used the fitted models to predict bycatch probability (binomial component of the delta-model) and density (positive component of the delta-model) at the test locations. This gave us predicted and observed values to compare parametric and non-parametric model performance: area under the receiver operating characteristic curve (AUC) scores for the binomial component and root-mean-square error (RMSE) for the positive component. AUC can be interpreted as the expectation that a model ranks a uniformly drawn random positive (bycatch event) as more likely than a uniformly drawn random negative (non-bycatch event), and higher values indicate better performance (Hand 2009). We tested for significance across and within species using Tukey’s HSD test.

Use as tools to reduce bycatch

Since AUC and RMSE are abstract metrics of model performance, we calculated a more tangible measure: how much each of the models could possibly reduce the bycatch-to-target species catch ratio if it were used to identify and remove high-bycatch-risk fishing events. In other words, what does an AUC performance gap of 0.03 mean in terms of bycatch reduction (keeping catch of the target species constant)? For each species and model fit, we rank-ordered the bycatch probabilities predicted by the binomial component, identified the X% of fishing events with the highest bycatch probabilities (X% from 0% to 10%), removed both bycatch and target catch from those events, and calculated the resultant change in bycatch-to-target species catch ratio.

To be useful bycatch management tools, models also need to generate reasonable spatial predictions and covariate effects and pass diagnostic checks on spatial autocorrelation and residual patterns. To investigate spatial predictions and their uncertainty, we calculated the posterior mean and variance at each point on a prediction grid, using the R package “mgcv” for GLMs and GAMs (Wood 2017) and “INLA” for GMRFs (Lindgren and Rue 2015). It is not possible to calculate posterior distributions for RF, but we calculated standard errors using the infinitesimal jackknife estimator (“randomForestCI”; Wager et al. 2014). To compare covariate effects, we plotted marginal posterior distributions for GLMs, GAMs, and GMRFs. To visualize RF covariate effects, we used feature contributions calculated by the “forestFloor” package (Welling et al. 2016). Lastly, we investigated the models’ abilities to

Fig. 2. Predictive performance boxplots of the (A) binomial and (B) positive components of the delta-model on test data from fivefold cross-validation repeated 10 times: (A) area under the receiver operating characteristic curve (AUC) for the binomial component and (B) normalized root-mean-square error (RMSE) for the positive component. Across species, random forests (RFs) outperformed generalized additive models (GAMs) and Gaussian Markov random fields (GMRFs) (highest AUC, lowest RMSE). Significant ($p < 0.05$, Tukey's HSD) within-species performance differences from RF and GMRF are denoted with black and blue asterisks, respectively. Only the best submodel (e.g., CONSTANT or YEAR) within each model class for each species is shown here (see online Supplementary material¹). Species abbreviations: DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye rockfish, LOGG = loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark. [Colour online.]

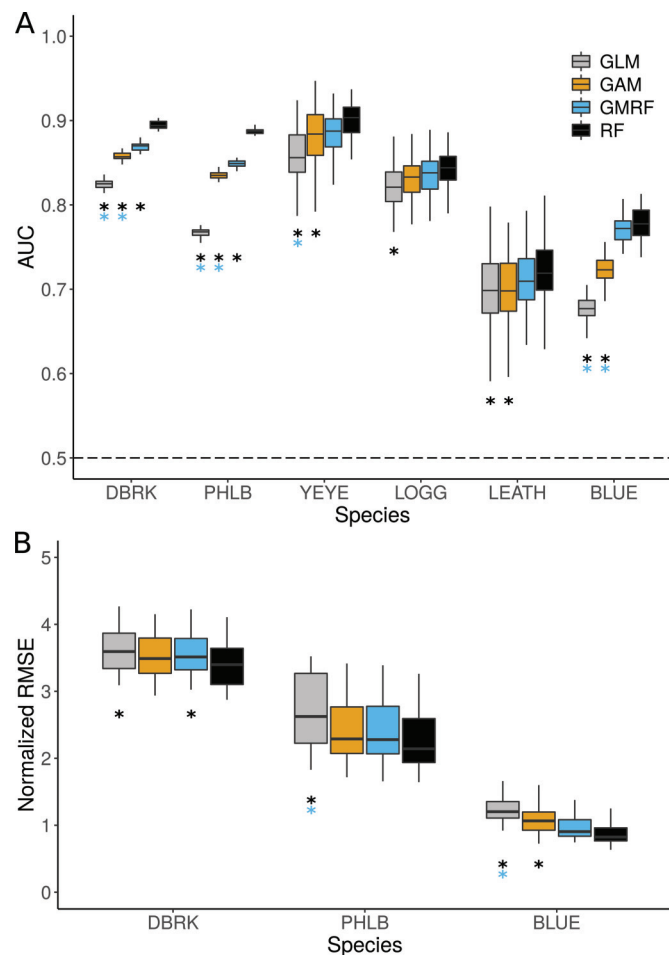
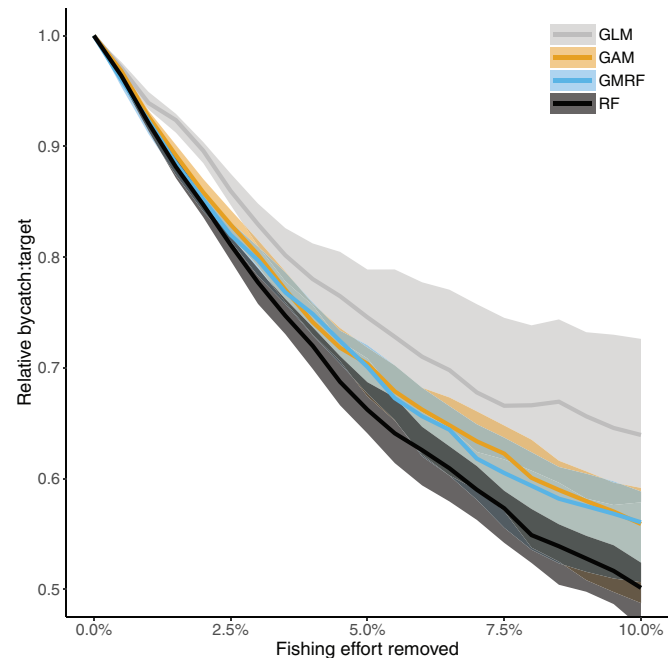


Fig. 3. Bycatch-to-target species catch ratio achieved by using the binomial component of the delta-model to predict and remove fishing sets in the test data, relative to the bycatch-to-target ratio with no fishing sets removed. Lines show median of 50 cross-validation runs for each model class (fivefold cross-validation repeated 10 times), averaged across the six species. Shaded areas are bootstrapped 95% confidence intervals for the median. Random forest (RF) performed the best, reducing the bycatch-to-target ratio by 34% when removing 5% of fishing and by 50% when removing 10% of fishing. As in Fig. 2, only the best submodel within each model class (e.g., CONSTANT or YEAR) for each species is shown here. [Colour online.]



0.5%, 1%, 2%, 5%, 10%, and 20% density locations as test data sets. We then fit the models using only the core, data-rich area of each fishery and computed AUC and RMSE for model predictions at the test locations in sparsely sampled areas.

Results

Performance metrics

Across the six species, RF provided better bycatch predictions than both GAMs and GMRFs in the binomial (higher AUC; Fig. 2A) and positive (lower RMSE; Fig. 2B) components of the delta-model. However, the magnitude of this performance advantage varied by species, and in some cases the within-species differences were not significant. GMRFs outperformed GAMs in the binomial component for the three species with moderate–high bycatch rates ($p < 0.05$, Tukey's HSD; darkblotched rockfish, Pacific halibut, and blue shark in Fig. 2A), but differences in AUC were not significant for the three rarely caught species ($p > 0.05$, Tukey's HSD; yelloweye rockfish, loggerhead turtle, and leatherback turtle in Fig. 2A). The variability in model performance among cross-validation runs was similar within a given species but varied greatly between species; variability in the binomial component was lowest for species with moderate bycatch rates (darkblotched rockfish and Pacific halibut in Fig. 2A), and in the positive component it was lowest for species with high bycatch rates (blue shark in Fig. 2B).

RF modifications designed for data with imbalanced classes (i.e., down-sampling and SMOTE) outperformed the original RF algorithm for the four species with high or low bycatch rates

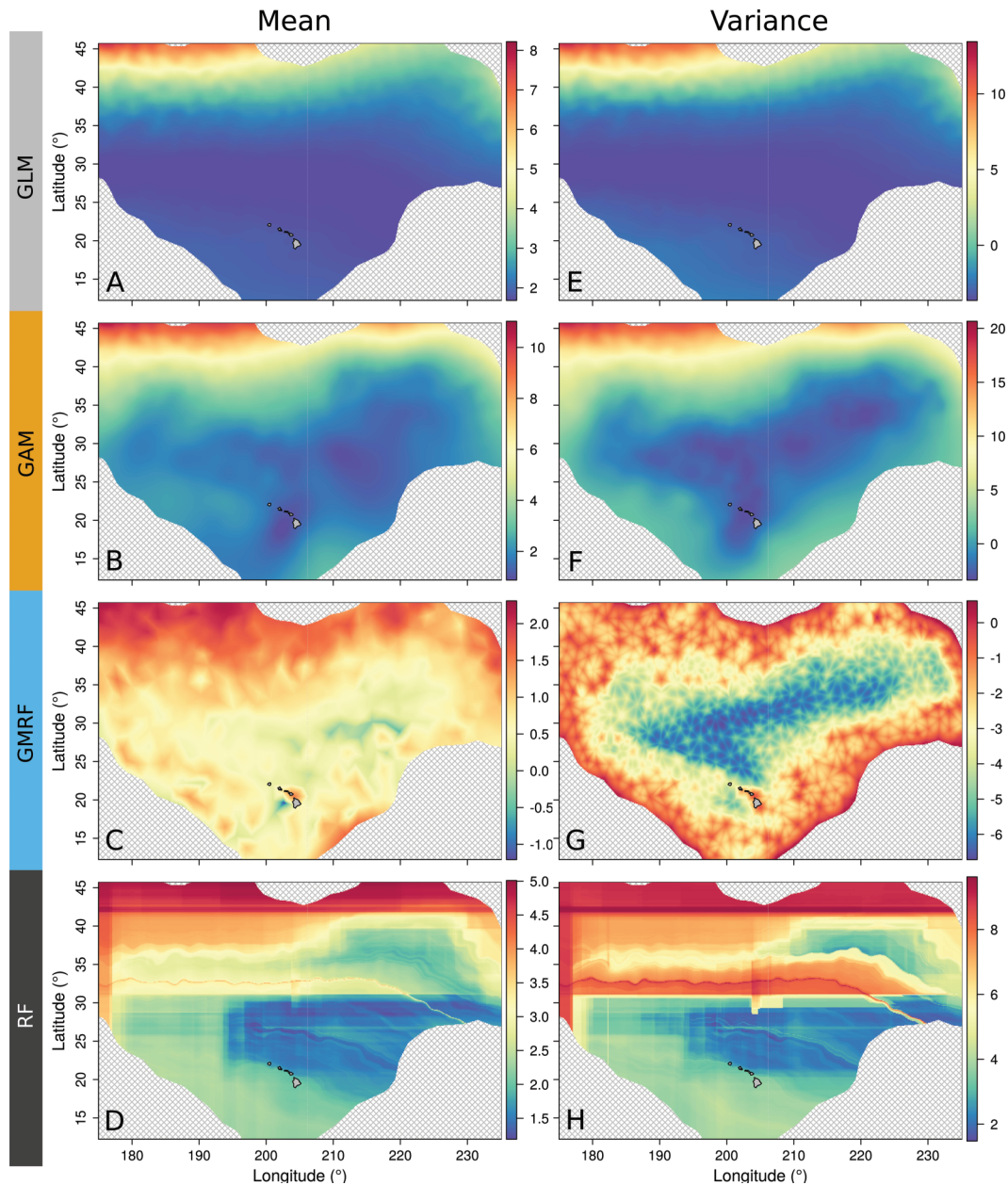
reduce spatial autocorrelation using spline correlograms ("nfc"; Bjornstad and Falck 2001) and binned residual plots.

Transferability

To evaluate model transferability (i.e., spatial extrapolation performance), we conducted a second cross-validation blocking on spatial data density. Whereas our first cross-validation procedure partitioned the data into training and testing sets randomly, here we constructed test–train splits by ordering fishing locations relative to data density. We calculated a bivariate kernel density estimate at each of the observed fishing locations ("bkde2D" function in "KernSmooth" R package) and sequentially used the lowest

¹Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2018-0281>.

Fig. 4. Maps of predicted blue shark bycatch density with uncertainty for the Hawaii longline swordfish fishery in 2014. Left panels show the mean bycatch density, $\log(\text{number per set})$, estimated using the four model frameworks: (A) GLM, (B) GAM, (C) GMRF, and (D) RF. Right panels show the log-variance of bycatch density: (E) GLM, (F) GAM, (G) GMRF, and (H) RF. All models predict higher and more variable bycatch density in the northwest area of the fishery. Maps created by GMRF and RF show artifacts of their construction; the mesh triangulation is evident in the GMRF variance map (G), and the sharp gradients in the RF mean (D) and variance (H) maps are a consequence of RF trees splitting on latitude and longitude. [Colour online.]



(Fig. A1). Of the semiparametric models, GMRF-YEAR that allowed for time-varying spatial effects performed the best (Figs. 2 and A2), whereas the binomial GAM-YEAR that estimated two-dimensional splines by year failed to converge. For rare bycatch species with few positive occurrences, such as yelloweye rockfish, GMRF models that allowed for time-varying spatial effects offered no improvement over the time-constant spatial models (Fig. A2). For these species, the estimated GMRF spatial field was less complex than for species with higher bycatch rates (Fig. A3).

Use as tools to reduce bycatch

When using the models to identify and remove high-bycatch-risk fishing events, RF also performed best. Averaged across the

six species, RF reduced the bycatch-to-target species catch ratio by 8% when removing 1% of fishing effort, 34% when removing 5% of fishing effort, and 50% when removing 10% of fishing effort (Fig. 3). Bycatch predictability as measured by bycatch-to-target ratio reduction generally agreed with the traditional performance metrics (AUC and RMSE) and varied substantially among the six species (Fig. A4).

Bycatch risk maps produced by each of the models were similar in some respects. Taking blue shark bycatch density an example, all models predicted higher and more variable bycatch in the northwest area of the fishery and lower bycatch between 25°N–30°N and 205°E–220°E (Figs. 4A–4D). Maps revealed artifacts of

Fig. 5. Covariate effects on the probability of darkblotched rockfish bycatch (binomial component of the delta-model). All models estimate a positive effect of PredOcc (predicted occurrence from survey data, left column), quadratic effect of Depth (center), and positive effect of In/near RCA (haul location inside or near rockfish conservation area boundary, right). GLM, GAM, and GMRF covariate effects are marginal posterior distributions (“mgcv” and “INLA” packages in R), and RF covariate effects are feature contributions (“forestFloor” package in R). [Colour online.]

their construction; the mesh triangulation is evident in the GMRF variance (Fig. 4G), and the sharp gradients in the RF mean (Fig. 4D) and variance (Fig. 4H) are a consequence of RF trees splitting on latitude and longitude. As expected, the uncertainty of GMRF predictions was lower in areas of high data density (compare Figs. 1B and 4G). Uncertainty in RF predictions did not follow this trend — RF variance was extremely high above 41.5°N and moderately high between 30°N and 33°N (Fig. 4H). However, this was consistent with the data, as there were few observed sets north of 41.5°N, and these had higher and more variable bycatch (Fig. A5).

All models estimated similar covariate effects, as demonstrated for darkblotched rockfish (Fig. 5). The main covariate effects were as expected; probability of bycatch increased with higher survey-predicted occurrence, increased for tows inside or near RCAs, and showed an optimal depth range of 100–250 fathoms (1 fathom = 1.8288 m).

For all species, residuals were typically more variable between cross-validation fits as model complexity increased (RF > GMRF > GAM > GLM; see online Supplementary material, Figs. S1–S18¹). Most models did not exhibit spatial patterns in residuals, although residuals in the positive component of the delta-model were larger and more variable for blue shark and Pacific halibut north of 40°N and 47°N, respectively (Figs. S1–S2, S11–S12¹). GMRFs and GAMs had similarly lower residual spatial autocorrelation compared with the baseline GLMs, with one exception where the GMRF reduced spatial autocorrelation more than the GAM (blue shark positive model; Fig. A6). RF generally had the lowest spatial autocorrelation, and it was negative instead of positive at short distances. The HILL species had greater decorrelation distances than the West Coast groundfish species (distance at which spatial autocorrelation goes to zero, 40 versus 5 km; Fig. A6).

Transferability

As expected, all models performed worse at spatial extrapolation compared with interpolation (i.e., worse at predicting locations outside the core area of the fishery when trained using observations in areas of highest data density; Fig. 6). GLMs and GAMs generally performed worse in this test than GMRFs and RFs. GMRFs had lower RMSE than RFs in the positive component of the delta-model, but there was no difference in AUC in the binomial component (Wilcoxon signed-rank test, $p = 0.91$ for $AUC_{\text{GMRF}} \neq AUC_{\text{RF}}$, $p = 0.003$ for $RMSE_{\text{GMRF}} \neq RMSE_{\text{RF}}$). Compared with GMRFs, RF performance was also more sensitive to withholding data at the edge of the fishery (Fig. 6). The degree to which this was true, however, differed widely between species. For instance, the performance of both models was stable for some species (e.g., darkblotched rockfish in binomial component, blue shark in positive component), indicating that both models captured relevant spatial environment–bycatch relationships in the core area of the fishery and that these relationships remained valid at the edge of the fishery.

Discussion

Our results demonstrate the clear potential of species distribution models to predict fishing activity with higher bycatch rates. While the models’ performance varied considerably, even the worst performer (GLM without latitude and longitude) achieved AUC from 0.68 to 0.86 (Fig. 2) and reduced bycatch-to-target catch ratios on average by 25% for a 5% reduction in fishing effort (Fig. 3). Random forests (RFs) performed the best, achieving cross-validated AUC above 0.89 for all three West Coast ground-

fish species (Fig. 2) and reducing bycatch-to-target ratios by 34% for a 5% reduction in fishing effort averaged across the six species (Fig. 3). When extrapolating beyond the geographic range of the data, however, RFs were more sensitive to which data were withheld and performed similar to, or worse than, GMRFs (Fig. 6). This is consistent with previous work documenting that more data-driven, complex SDMs can have better interpolation performance but be worse at spatial extrapolation (Heikkinen et al. 2012; Randin et al. 2006).

Our results beg the question: if RFs are expected to have higher interpolation accuracy, why ever use GMRFs or GAMs? This is a decidedly relevant concern given the substantial research investment in these modeling approaches (Becker et al. 2014; Thorson and Barnett 2017). RF will generally have better out-of-sample prediction, largely due to its ability to incorporate nonlinear and interaction effects of covariates inherently (i.e., without user specification). In addition, RF models are much simpler and quicker to both write code for and run. Contrary to references describing RFs as “black boxes” (Prasad et al. 2006; Cutler et al. 2007; Elith and Leathwick 2009; Evans and Cushman 2009; Kuhn and Johnson 2013), there are methods for investigating RF model structure, including covariate effects and interactions (Fig. 5; Welling et al. 2016). In a similar vein, it is possible to calculate prediction variance and confidence intervals for RF (Fig. 4; Meinshausen 2006; Wager et al. 2014), despite older ecological literature stating otherwise (Cutler et al. 2007; Olden et al. 2008), or using ad hoc substitutes such as the standard deviation of individual tree predictions (Smoliński and Radtke 2017). Additionally, promising theoretical work may soon widen the ability to use RFs for statistical inference by developing asymptotically normal, unbiased point estimates with valid confidence intervals (Mentch and Hooker 2017; Wager and Athey 2017). In many cases, RF’s performance advantage is probably sufficient to warrant its use over other semiparametric methods. Yet, there may be several cases where semiparametric methods are preferred.

Semiparametric frameworks like GMRFs and GAMs have clear advantages over machine learning algorithms such as RF. First, they can be derived from probability theory and therefore allow for traditional statistical inference on their mean response predictions. GMRF estimates full posterior distributions for the response variable (in this study, probability and (or) expected amount of bycatch) everywhere in the spatial domain. This enables us, for instance, to use a GMRF model to identify regions that are above or below a threshold probability (i.e., risk level) of a defined bycatch quantity. By contrast, developing statistical inference based on RF is an active area of research (Biau and Scornet 2016). Thus, GMRF models may be preferred for applications where estimates of model uncertainty are decidedly important, such as using models to produce annual estimates of bycatch (expanded from the observed to unobserved fleet). Second, GMRFs and GAMs explicitly estimate covariate effects with uncertainty intervals, facilitating ecological interpretation of factors significantly affecting the response (Fig. 5). Third, GMRF models are particularly well suited to incorporate spatial ecological processes, such as movement or spatial variation in mortality, condition, or observation error (Carson and Flemming 2014; Illian et al. 2013; Thorson et al. 2017). RFs, on the other hand, are based solely on observations without taking into consideration the data-generating process. Fourth, GAMs and GMRFs produce smoothed spatial predictions that are more likely ecologically plausible, whereas RF spatial predictions appear rectangular by default due to splits on the geographical

Can. J. Fish. Aquat. Sci. Downloaded from cdsciencepub.com by NOAA CENTRAL on 06/05/23
For personal use only.

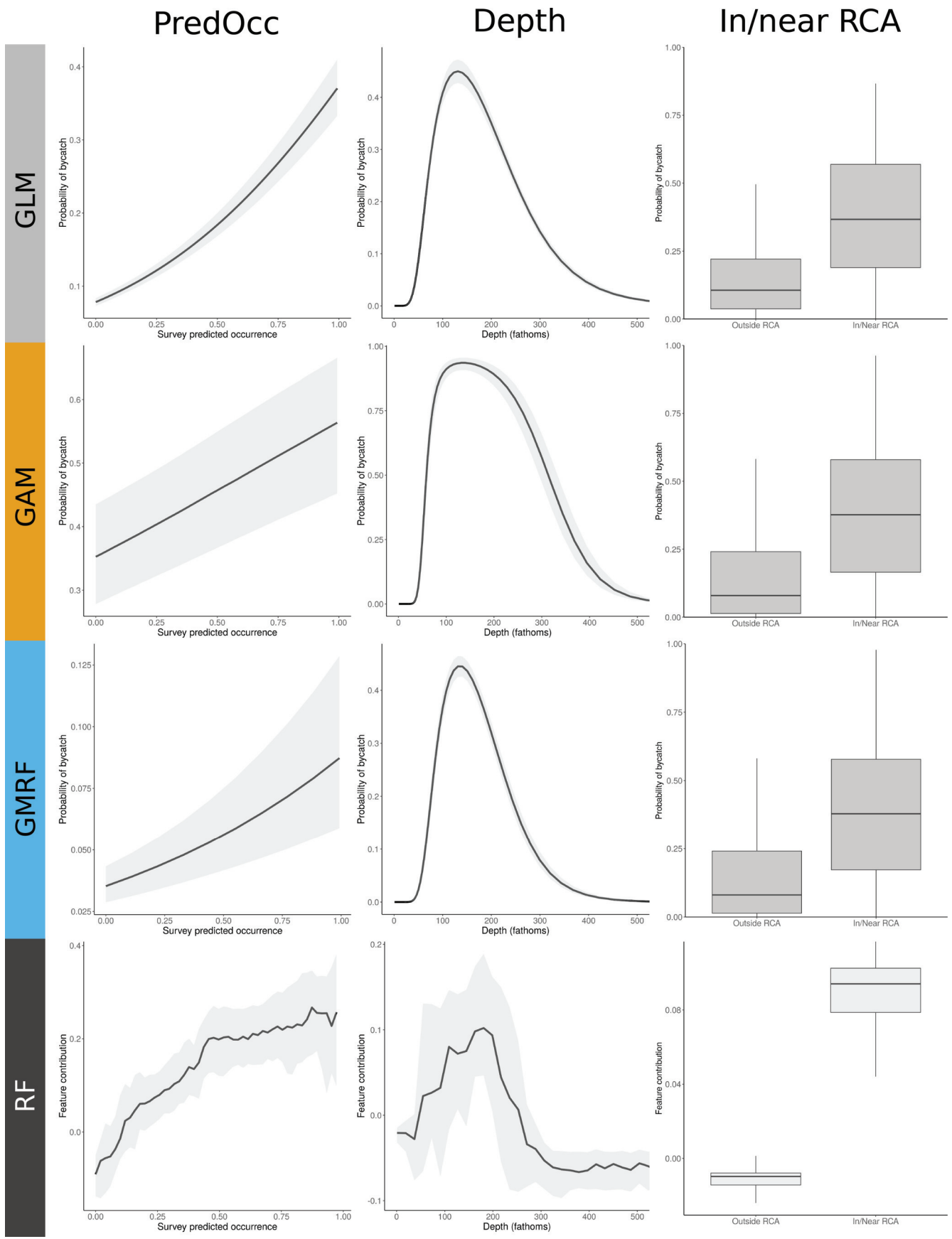
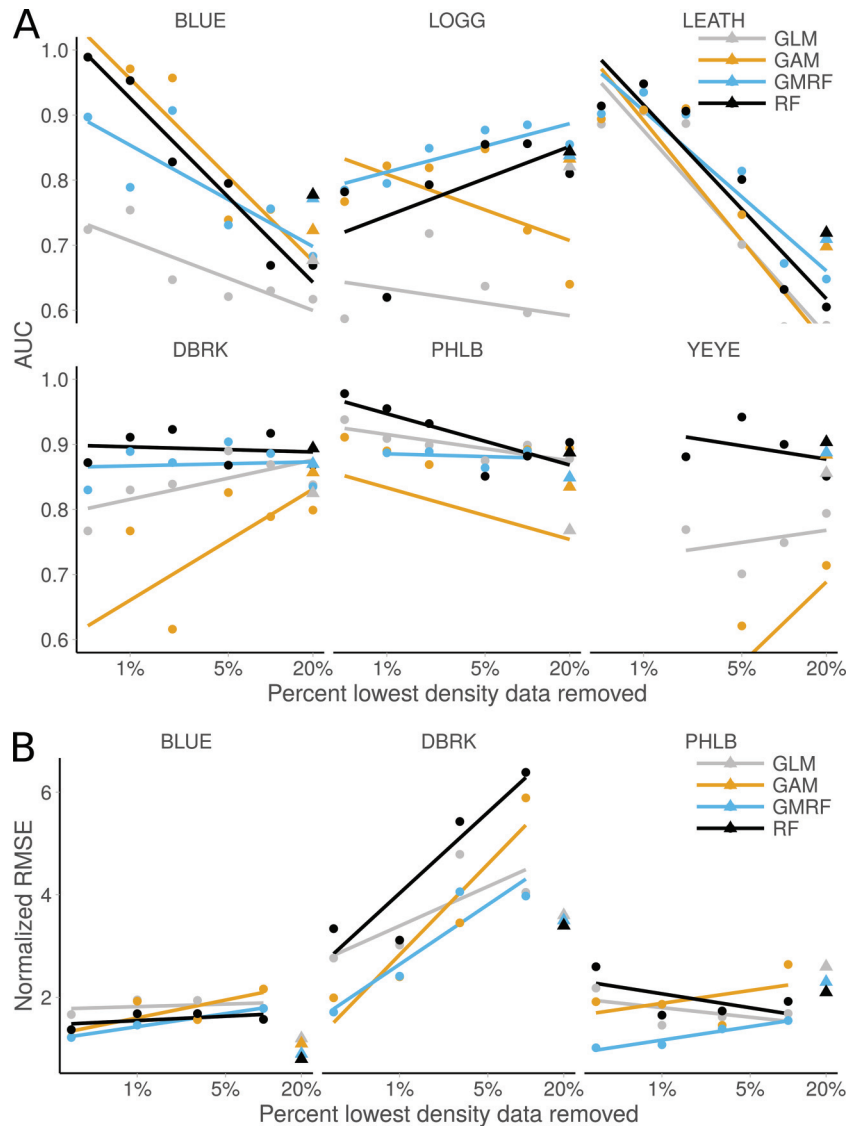


Fig. 6. Predictive performance of the (A) binomial and (B) positive components of the delta-models at test locations beyond the geographic data range (i.e., spatial extrapolation). We fit a two-dimensional kernel density estimate at each observed fishing location (“bkde2D” function in “KernSmooth” R package), then sequentially used the lowest 0.5%, 1%, 2%, 5%, 10%, and 20% density locations as test data sets. Triangles show median model performance from fivefold cross-validation runs with random test–train splits (Fig. 2). When extrapolating spatially, all models performed equal to or worse than when interpolating (i.e., points are lower AUC and higher RMSE than triangles at 20% removed). Compared with GMRF, RF performance was more sensitive to withholding data at the edge of the fishery (i.e., regression lines have steeper slopes). Missing points and lines indicate the model failed to converge, as for GMRF with yelloweye rockfish. Species abbreviations: DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye rockfish, LOGG = loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark. [Colour online.]

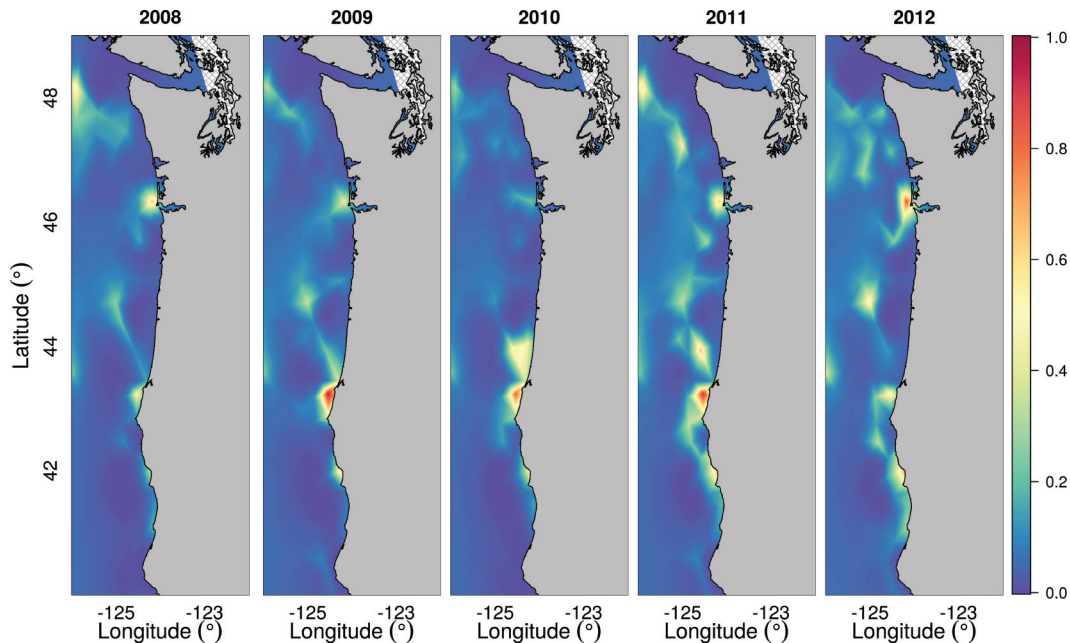


coordinates. This can be mediated, however, by including transformations of the coordinates, fitting a linear model within each node instead of taking the mean (Quinlan 1992, 1993), or including buffer distances as covariates (Hengl et al. 2018). Fifth, RFs can simultaneously have high interpolation accuracy and lower extrapolation accuracy (Fig. 6B). This matters if SDM predictions of bycatch risk are to be used as a spatial management tool, where predictions in areas with sparse or no sampling coverage may be most important. Finally, GMRFs and GAMs can be easily specified to produce independent or exchangeable estimates of a given quantity (e.g., total predicted bycatch in different years), while it is unclear how to assign a specific dependence structure on factors in a RF model. This is important when using estimates from a spatial model as input in a secondary model, for example, where the secondary model assumes that estimates are independent

among years. In situations where these concerns are inconsequential, however, RFs are likely the better method for spatial bycatch prediction — they are faster and have better predictive performance than the alternatives. Even when probabilistic conclusions are required, RF may be useful in an exploratory manner given their ease of use, speed, and ability to identify nonlinear and interaction covariate effects for later inclusion in semiparametric models.

On a more detailed level, the six species differed widely in their predictability. The West Coast groundfish species were more predictable than the HILL species, likely because we included several more relevant environmental covariates (WCGOP: SST, depth, rocky habitat, in or near RCA, and survey-predicted occurrence; HILL: SST), and the WCGOP data set had 2.5 times the number of observations (WCGOP: 42 786; HILL: 16 714). The HILL models' per-

Fig. 7. GMRF-YEAR random field for bycatch probability of darkblotched rockfish from 2008 to 2012. [Colour online.]



formance could presumably be improved by incorporating more satellite-based environmental covariates capable of explaining the species' distributions, such as chlorophyll and SST-derived frontal indices (Nieto et al. 2017). Among the West Coast species, Pacific halibut was more difficult to predict than the two rockfish species, despite having more positive bycatch occurrences for the models to fit. One possible explanation is that adult halibut move more and have less strict habitat associations than rockfish, decreasing their predictability (Skud 1977; Gunderson 1997). Among the HILL species, loggerhead turtles were much easier to predict than leatherback turtles and blue sharks, perhaps because they have stronger SST-based habitat preferences where the fishing occurs (Howell et al. 2015). Indeed, the loggerhead-temperature association is the basis for TurtleWatch, a decade-long effort to reduce turtle bycatch by providing fishermen with dynamic recommendations of high bycatch risk areas to avoid in the Central Pacific (Howell et al. 2008). Despite plausible life history explanations for some among-species differences in predictability, however, the advantage of fitting time-varying versus time-constant models appeared to be primarily driven by the species' bycatch rates. In other words, there was little difference in model performance for rare bycatch species (yelloweye rockfish, loggerhead turtle, and leatherback turtle) and greater differences between models for common bycatch species (darkblotched rockfish, Pacific halibut, and blue shark; Figs. 2A and A3).

Given their good predictive performance overall, SDMs could be used to support spatial bycatch management, whether static (e.g., design habitat closures to be semipermanent, such as the Pacific Leatherback Conservation Area along the US West Coast; 50 CFR Part 660) or dynamic in time (e.g., closures change every year, month, week, etc., such as the Loggerhead Turtle Conservation Area along the Southern California Bight; 72FR 31756; Fig. 7, as in Dunn et al. 2016). If so, they should be compared with and integrated with existing tools that aim to reduce bycatch by producing risk maps; examples include Eguchi et al. (2017), TurtleWatch (Howell et al. 2008, 2015), and WhaleWatch (Hazen et al. 2016; Breivik et al. 2016). Both TurtleWatch and WhaleWatch are based on satellite telemetry observations and known habitat preferences of sea turtles and blue whales, in contrast with the models developed here that rely exclusively on fisheries observer data. Fisheries observer data sets cover many more bycatch species

than those with satellite tagging programs, which means that SDMs based on fisheries observer data may be more widely applicable than those based solely on satellite telemetry, especially for species with moderate to high bycatch rates. However, we found that SDMs were less effective at predicting rare bycatch events (e.g., yelloweye rockfish, loggerhead turtle, and leatherback turtle in Fig. 2A). Bycatch occurs when nontarget species and fishing gear co-occur, both of which are affected by various factors, such as environmental conditions, economics, and behavior (Soykan et al. 2014). Consequently, using fishery observer data combined with animal movement data would provide a comprehensive data set to develop predictive models of bycatch, and this may be a prudent approach for species with low bycatch rates.

Future efforts to use spatial models to predict fisheries bycatch risk should carefully consider the hierarchical structure common to observer data sets with less than 100% coverage. In the typical case where observers are placed on vessels on a trip-by-trip basis and then observe all sets within a trip, the sets are likely not independent (e.g., sets within the same trip, and trips on the same vessel may be correlated). As done in this study, including spatio-temporal correlation structure will account for some of the correlation between sets within a trip because they are presumably closer together in time and space. One approach is to include the nested data structure in the model as random effects (Candy 2004; Thorson et al. 2015b), although this can be more complicated than it first appears (e.g., captains and crew transfer between vessels in the HILL fishery, making it unclear whether including a vessel effect is appropriate). Roberts et al. (2017) make an excellent case for an alternative approach, using block cross-validation to account for spatial, temporal, and group dependence structure when validating and selecting models. A related issue with the delta-models used here is that they assume two independent processes determine the probability and amount of bycatch (i.e., the binomial and positive components of the delta-model). Thorson (2018) and Cantoni et al. (2017) both recently demonstrated that this is unlikely to be true and proposed solutions that include parameters allowing for dependence between the binomial and positive components. Another relatively simple approach is to use the compound Poisson-gamma, or Tweedie distribution, which outperformed the equivalent delta-GAM in a recent spatial by-

catch analysis (Stock et al. 2019). All of these solutions should improve spatiotemporal bycatch predictions in the future.

Finally, just as single-species fisheries management paints a rosier picture than can truly be implemented, the results presented here are unrealistic in their treatment of multispecies fisheries because bycatch prediction cannot be optimized for each individual species simultaneously. These results are still useful if fisheries managers are particularly concerned about a single species, but less so if reducing bycatch of multiple species is the objective. Both RF and GMRF models have multivariate extensions that could fruitfully be applied to multispecies spatial bycatch prediction, and future work should investigate this possibility (Thorson et al. 2015a; Ishwaran and Kogalur 2018; Thorson and Barnett 2017).

Data accessibility

Unsummarized fisheries observer data are deemed confidential. Code to download and process publicly available fisheries survey data, run each of the models, and replicate figures are provided at <https://github.com/brianstock/spatial-bycatch>. Summarized reports of the fisheries observer program data are available at https://www.nwfsc.noaa.gov/research/divisions/fram/observation/data_products/data_library.cfm (US West Coast groundfish) and http://www.fpir.noaa.gov/OBS/obs_hi_ll_ss_rprts.html (Hawaii longline).

Author contributions

BCS, EW, TE, and BXS conceived of and designed the study with input from JJ, JT, and BF; BCS, JJ, and BF curated the data; BCS performed the analysis with input from EW and BXS; BCS, EW, and JT drafted the manuscript; all authors provided critical reviews.

Acknowledgements

T. Todd Jones, Summer Martin, Don Kobayashi, Eric Forney, and Marti McCracken shared valuable knowledge of the Hawaii longline fishery and observer program. BCS received support from the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144086. BCS also received support from the NMFS Protected Species Toolbox and benefited from discussions at two NMFS Protected Species Assessment Workshops. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation grant No. ACI-1053575.

References

Alverson, D.L., Freeberg, M.H., Murawski, S.A., and Pope, J.G. 1994. A global assessment of fisheries bycatch and discards. FAO Fisheries Technical Paper. No. 339. FAO, Rome.

Aratijo, M.B., and Rahbek, C. 2006. How does climate change affect biodiversity? *Science*, **313**: 1396–1397. doi:10.1126/science.1131758. PMID:16959994.

Becker, E.A., Forney, K.A., Foley, D.G., Smith, R.C., Moore, T.J., and Barlow, J. 2014. Predicting seasonal density patterns of California cetaceans based on habitat models. *Endang. Species Res.* **23**: 1–22. doi:10.3354/esr00548.

Bellman, M.A., Heery, E., and Majewski, J. 2010. Observed and estimated total bycatch of green sturgeon in the 2002–2008 U.S. West Coast groundfish fisheries. West Coast Groundfish Observer Program. NWFSC, 2725 Montlake Blvd E., Seattle, WA 98112, USA.

Benson, S.R., Eguchi, T., Foley, D.G., Forney, K.A., Bailey, H., Hitipeuw, C., et al. 2011. Large-scale movements and high-use areas of western Pacific leatherback turtles, *Dermochelys coriacea*. *Ecosphere*, **2**(7): 1–27. doi:10.1890/ES11-00053.1.

Biau, G., and Scornet, E. 2016. A random forest guided tour. *TEST*, **25**: 197–227. doi:10.1007/s11749-016-0481-7.

Bjornstad, O.N., and Falck, W. 2001. Nonparametric spatial covariance functions: estimation and testing. *Environ. Ecol. Stat.* **8**: 53–70. doi:10.1023/A:1009601932481.

Boyce, J.R. 1996. An economic analysis of the fisheries bycatch problem. *J. Environ. Econ. Manage.* **31**(3): 314–336. doi:10.1006/jeeem.1996.0047.

Bradburn, M.J., Keller, A.A., and Horness, B.H. 2011. The 2003 to 2008 U.S. West Coast bottom trawl surveys of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, length, and age composition. US Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-114.

Breiman, L. 2001. Random forests. *Mach. Learn.* **45**: 5–32. doi:10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R.A., and Stone, C.J. 1984. Classification and regression trees. Chapman & Hall/CRC, Boca Raton, Fla.

Breivik, O.N., Storvik, G., and Nedreaas, K. 2016. Latent Gaussian models to decide on spatial closures for bycatch management in the Barents Sea shrimp fishery. *Can. J. Fish. Aquat. Sci.* **73**: 1271–1280. doi:10.1139/cjfas-2015-0322.

Breivik, O.N., Storvik, G., and Nedreaas, K. 2017. Latent Gaussian models to predict historical bycatch in commercial fishery. *Fish. Res.* **185**: 62–72. doi:10.1016/j.fishres.2016.09.033.

Candy, S.G. 2004. Modelling catch and effort data using generalised linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Sci.* **11**: 59–80.

Cantoni, E., Flemming, J.M., and Welsh, A.H. 2017. A random-effects hurdle model for predicting bycatch of endangered marine species. *Ann. Appl. Stat.* **11**(4): 2178–2199. doi:10.1214/17-AOAS1074.

Carretta, J.V., Moore, J.E., and Forney, K.A. 2017. Regression tree and ratio estimates of marine mammal, sea turtle, and seabird bycatch in the California drift gillnet fishery: 1990–2015. NOAA Tech. Memo., NOAA-TM-NMFS-SWFSC-568.

Carson, S., and Flemming, J.M. 2014. Seal encounters at sea: A contemporary spatial approach using R-INLA. *Ecol. Modell.* **291**: 175–181. doi:10.1016/j.ecolmodel.2014.07.022.

Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P. 2002. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**: 321–357. doi:10.1613/jair.953.

Conn, P.B., Johnson, D.S., Ver Hoef, J.M., Hooten, M.B., London, J.M., and Boveng, P.L. 2015. Using spatiotemporal hierarchical models to estimate animal abundance and infer ecological dynamics from survey counts. *Ecol. Monogr.* **85**(2): 235–252. doi:10.1890/14-0959.1.

Cosandey-Godin, A., Krainski, E.T., Worm, B., and Flemming, J.M. 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. *Can. J. Fish. Aquat. Sci.* **72**: 186–197. doi:10.1139/cjfas-2014-0159.

Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. 2007. Random forests for classification in ecology. *Ecology*, **88**: 2783–2792. doi:10.1890/07-0539.1. PMID:18051647.

Davies, R.W.D., Cripps, S.J., Nickson, A., and Porter, G. 2009. Defining and estimating global marine fisheries bycatch. *Mar. Policy*, **33**(4): 661–672. doi:10.1016/j.marpol.2009.01.003.

Diggle, P.J., Tawn, J.A., and Moyeed, R.A. 1998. Model-based geostatistics. *J. R. Stat. Soc. Ser. C, Appl. Stat.* **47**(3): 299–350. doi:10.1111/1467-9876.00113.

Drake, J.M., Randin, C., and Guisan, A. 2006. Modelling ecological niches with support vector machines. *J. Appl. Ecol.* **43**: 424–432. doi:10.1111/j.1365-2664.2006.01141.x.

Dunn, D.C., Maxwell, S.M., Boustany, A.M., and Halpin, P.N. 2016. Dynamic ocean management increases the efficiency and efficacy of fisheries management. *Proc. Natl. Acad. Sci.* **113**: 668–673. doi:10.1073/pnas.1513626113. PMID:26729885.

Eguchi, T., Benson, S.R., Foley, D.G., and Forney, K.A. 2017. Predicting overlap between drift gillnet fishing and leatherback turtle habitat in the California Current Ecosystem. *Fish. Oceanogr.* **26**(1): 17–33. doi:10.1111/fog.12181.

Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* **40**: 677–697. doi:10.1146/annurev.ecolsys.110308.120159.

Evans, J.S., and Cushman, S.A. 2009. Gradient modeling of conifer species using random forests. *Landsc. Ecol.* **24**(5): 673–683. doi:10.1007/s10980-009-9341-0.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. 2013. Nonparametric regression. *In* Regression: models, methods and applications. Springer-Verlag, Berlin, Heidelberg. pp. 413–533.

FAO. 1995. Code of Conduct for Responsible Fisheries. Rome, FAO.

Gilman, E., Kobayashi, D., Swenarton, T., Brothers, N., Dalzell, P., and Kinan-Kelly, I. 2007. Reducing sea turtle interactions in the Hawaii-based longline swordfish fishery. *Biol. Conserv.* **139**: 19–28. doi:10.1016/j.biocon.2007.06.002.

Golding, N., and Purse, B.V. 2016. Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods Ecol. Evol.* **7**(5): 598–608. doi:10.1111/2041-210X.12523.

Guélat, J., and Kéry, M. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods Ecol. Evol.* **9**(6): 1614–1625. doi:10.1111/2041-210X.12983.

Guisan, A., and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**: 993–1009. doi:10.1111/j.1461-0248.2005.00792.x.

Gunderson, D.R. 1997. Spatial patterns in the dynamics of slope rockfish stocks and their implications for management. *Fish. Bull.* **95**: 219–230.

Hall, M.A. 1999. Estimating the ecological impacts of fisheries: what data are needed to estimate bycatches? *In* Proceedings of the International Conference on Integrated Fisheries Monitoring. Edited by C.P. Nolan. Paper presented at The International Conference on Integrated Fisheries Monitoring, Sydney, Australia, 1–5 February. Rome, FAO. pp. 175–184.

Hand, D.J. 2009. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach. Learn.* **77**(1): 103–123. doi:10.1007/s10994-009-5119-5.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The elements of statistical learning. 2nd edition. Springer New York.

- Hazen, E.L., Palacios, D.M., Forney, K.A., Howell, E.A., Becker, E., Hoover, A.L., et al. 2016. WhaleWatch: a dynamic management tool for predicting blue whale density in the California Current. *J. Appl. Ecol.* **54**(5): 1415–1428. doi:10.1111/1365-2664.12820.
- Heikkinen, R.K., Marmion, M., and Luoto, M. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, **35**(3): 276–288. doi:10.1111/j.1600-0587.2011.06999.x.
- Hengl, T., Nussbaum, M., Wright, M.N., and Heuvelink, G.B. 2018. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ Preprints* e26693v1.
- Hooten, M.B., and Hobbs, N.T. 2015. A guide to Bayesian model selection for ecologists. *Ecology*, **85**: 3–28. doi:10.1890/14-0661.1.
- Howell, E.A., Kobayashi, D.R., Parker, D.M., Balazs, G.H., and Polovina, J.J. 2008. TurtleWatch: a tool to aid in the bycatch reduction of loggerhead turtles *Caretta caretta* in the Hawaii-based pelagic longline fishery. *Endang. Species Res.* **5**: 267–278. doi:10.3354/esr00096.
- Howell, E.A., Hoover, A., Benson, S.R., Bailey, H., Polovina, J.J., Seminoff, J.A., and Dutton, P.H. 2015. Enhancing the TurtleWatch product for leatherback sea turtles, a dynamic habitat model for ecosystem-based management. *Fish. Oceanogr.* **24**: 57–68. doi:10.1111/fog.12092.
- Ianelli, J.N., and Stram, D.L. 2015. Estimating impacts of the pollock fishery bycatch on western Alaska Chinook salmon. *ICES J. Mar. Sci.* **72**: 1159–1172. doi:10.1093/icesjms/fsu173.
- Illian, J.B., Martino, S., Sørbye, S.H., Gallego-Fernández, J.B., Zunzunegui, M., Esquivias, M.P., and Travis, J.M.J. 2013. Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods Ecol. Evol.* **4**: 305–315. doi:10.1111/2041-210x.12017.
- Ishwaran, H., and Kogalur, U.B. 2018. Random forests for survival, regression and classification (RF-SRC). R package version 2.6.1.
- Kammann, E.E., and Wand, M.P. 2003. Geoadaptive models. *J. R. Stat. Soc.: Ser. C, Appl. Stat.* **52**(1): 1–18. doi:10.1111/1467-9876.00385.
- Kelleher, K. 2005. Discards in the world's marine fisheries: An update. FAO Fisheries Technical Paper. No. 470. Rome, FAO.
- Kneib, T., Müller, J., and Hothorn, T. 2008. Spatial smoothing techniques for the assessment of habitat suitability. *Environ. Ecol. Stat.* **15**(3): 343–364. doi:10.1007/s10651-008-0092-x.
- Kobayashi, D.R., Polovina, J.J., Parker, D.M., Kamezaki, N., Cheng, I.J., Uchida, I., et al. 2008. Pelagic habitat characterization of loggerhead sea turtles, *Caretta caretta*, in the North Pacific Ocean (1997–2006): insights from satellite tag tracking and remotely sensed data. *J. Exp. Mar. Biol. Ecol.* **356**(1): 96–114. doi:10.1016/j.jembe.2007.12.019.
- Kuhn, M., and Johnson, K. 2013. Applied predictive modeling. Springer Science & Business Media, New York.
- Leathwick, J.R., Elish, J., and Hastie, T. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Modell.* **199**: 188–196. doi:10.1016/j.ecolmodel.2006.05.022.
- Lecomte, J.B., Benoit, H.P., Ancelet, S., Etienne, M.P., Bel, L., and Parent, E. 2013. Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. *Methods Ecol. Evol.* **4**(12): 1159–1166. doi:10.1111/2041-210X.12122.
- Li, S., and Pan, M. 2011. Fishing opportunities under the sea turtle interaction caps—a spatial bio-economic model for Hawaii-based longline swordfish fishery. SOEST 11-02. Joint Institute for Marine and Atmospheric Research (JIMAR) Contribution 11-378.
- Liaw, A., and Wiener, M. 2002. Classification and regression by random forest. *R News*, **2**: 18–22.
- Liggins, G.W., Bradley, M.J., and Kennelly, S.J. 1997. Detection of bias in observer-based estimates of retained and discarded catches from a multi species trawl fishery. *Fish. Res.* **32**(2): 133–147. doi:10.1016/S0165-7836(97)00053-2.
- Lindgren, F., and Rue, H. 2015. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63**: 1–25.
- 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. *J. R. Stat. Soc.: Ser. B, Stat. Methodol.* **73**: 423–498. doi:10.1111/j.1467-9868.2011.00777.x.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., and Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modeling. *Divers. Distrib.* **15**(1): 59–69. doi:10.1111/j.1472-4642.2008.00491.x.
- Martin, S.L., Stohs, S.M., and Moore, J.E. 2015. Bayesian inference and assessment for rare-event bycatch in marine fisheries: a drift gillnet fishery case study. *Ecol. Appl.* **25**(2): 416–429. doi:10.1890/14-0059.1. PMID:26263664.
- Mauder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* **70**(2): 141–159. doi:10.1016/j.fishres.2004.08.002.
- McCracken, M.L. 2004. Modeling a very rare event to estimate sea turtle bycatch: Lessons learned. US Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-3.
- Meinshausen, N. 2006. Quantile regression forests. *Journal of Machine Learning Research*, **7**: 983–999.
- Mentch, L., and Hooker, G. 2017. Formal hypothesis tests for additive structure in random forests. *J. Comput. Graph. Stat.* **26**(3): 589–597. doi:10.1080/10618600.2016.1256817. PMID:30906174.
- Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., et al. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography*, **37**(12): 1267–1281. doi:10.1111/ecog.00845.
- Nichols, W.J., Resendiz, A., Seminoff, J.A., and Resendiz, B. 2000. Transpecific migration of a loggerhead turtle monitored by satellite telemetry. *Bull. Mar. Sci.* **67**(3): 937–947.
- Nieto, K., Xu, Y., Teo, S.L.H., McClatchie, S., and Holmes, J. 2017. How important are coastal fronts to albacore tuna (*Thunnus alalunga*) habitat in the Northeast Pacific Ocean? *Prog. Oceanogr.* **150**: 62–71. doi:10.1016/j.pocean.2015.05.004.
- NMFS. 2013. Groundfish essential fish habitat synthesis: a report to the Pacific Fishery Management Council. National Marine Fisheries Service, Seattle, Wash.
- NMFS. 2016. U.S. National Bycatch Report First Edition Update 2. Edited by L.R. Benaka, D. Bullock, J. Davis, E.E. Seney, and H. Winarsoo. US Dep. Commer., National Marine Fisheries Service.
- NOAA Fisheries West Coast Region. 2015. Rockfish conservation areas [online]. Available from http://www.westcoast.fisheries.noaa.gov/fisheries/management/groundfish_closures/rockfish_areas.html.
- NOAA National Centers for Environmental Information. 2015. U.S. Coastal Relief Model [online]. Available from <https://www.ngdc.noaa.gov/mgg/coastal/crm.html>.
- NWFSC. 2006. Observer coverage plan: Sampling plan and logistics for the West Coast Groundfish Observer Program [online]. Northwest Fisheries Science Center, NOAA, West Coast Groundfish Observer Program, 2725 Montlake Blvd. E, Seattle, Wash. Available from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.231.1675>.
- NWFSC. 2016. West Coast Groundfish Observer Program 2016 catch share training manual. Northwest Fisheries Science Center, Seattle, Wash.
- Olden, J.D., Lawler, J.J., and Poff, N.L. 2008. Machine learning methods without tears: a primer for ecologists. *Quart. Rev. Biol.* **83**: 171–193. doi:10.1086/587826. PMID:18605534.
- Parmesan, C., and Yohe, G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**: 37–42. doi:10.1038/nature01286. PMID:12511946.
- Pennington, M. 1983. Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, **39**: 281–286. doi:10.2307/2530830.
- Péron, G., Ferrand, Y., Gossman, F., Bastat, C., Guenezan, M., and Gimenez, O. 2011. Nonparametric spatial regression of survival probability: visualization of population sinks in Eurasian Woodcock. *Ecography*, **92**(8): 1672–1679. doi:10.1890/10-2224.1. PMID:21905433.
- Phillips, S.J., and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**: 161–175. doi:10.1111/j.0906-7590.2008.5203.x.
- Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* **190**: 231–259. doi:10.1016/j.ecolmodel.2005.03.026.
- 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. doi:10.1126/science.1239352. PMID:24031017.
- PIROF. 2014. Hawaii longline observer program observer field manual. Version LM.14.04. Pacific Islands Regional Observer Program, NOAA/IRC 1, 1845 Wasp Blvd. Bldg. 176, Honolulu, Hawaii 96818, USA.
- Pons, M., Marroni, S., Machado, I., Ghattas, B., and Domingo, A. 2009. Machine learning procedures: an application to by-catch data of the marine turtles *Caretta caretta* in the Southwestern Atlantic Ocean. *Collect. Vol. Sci. Pap. ICCAT*. **64**(7): 2443–2454.
- Prasad, A.M., Iverson, L.R., and Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**(2): 181–199. doi:10.1007/s10021-005-0054-1.
- Quinlan, J.R. 1992. Learning with continuous classes. In 5th Australian Joint Conference on Artificial Intelligence. pp. 343–348.
- Quinlan, J.R. 1993. Combining instance-based and model-based learning. In Proceedings of the Tenth International Conference on Machine Learning. pp. 236–243.
- R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., and Guisan, A. 2006. Are niche-based species distribution models transferable in space? *J. Biogeogr.* **33**(10): 1689–1703. doi:10.1111/j.1365-2699.2006.01466.x.
- Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., and Schlax, M.G. 2007. Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* **20**: 5473–5496. doi:10.1175/2007JCLI1824.1.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elish, J., Guillera-Aroita, G., et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **40**(8): 913–929. doi:10.1111/ecog.02881.
- Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc.: Ser. B, Stat. Methodol.* **71**: 319–392. doi:10.1111/j.1467-9868.2008.00700.x.
- Scales, K.L., Miller, P.I., Ingram, S.N., Hazen, E.L., Bograd, S.J., and Phillips, R.A. 2016. Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Divers. Distrib.* **22**: 212–224. doi:10.1111/ddi.12389.

- Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. *Can. J. Fish. Aquat. Sci.* **71**: 1655–1666. doi:10.1139/cjfas-2013-0508.
- Shmueli, G. 2010. To explain or to predict? *Stat. Sci.* **25**: 289–310. doi:10.1214/10-STS330.
- Skud, B.E. 1977. Drift, migration, and intermingling of Pacific halibut stocks. Scientific Report No. 63. International Pacific Halibut Commission, Seattle, Wash.
- Smoliński, S., and Radtke, K. 2017. Spatial prediction of demersal fish diversity in the Baltic Sea: comparison of machine learning and regression-based techniques. *ICES J. Mar. Sci.* **74**: 102–111. doi:10.1093/icesjms/fsw136.
- Soykan, C.U., Eguchi, T., Kohin, S., and Dewar, H. 2014. Prediction of fishing effort distributions using boosted regression trees. *Ecol. Appl.* **24**(1): 71–83. doi:10.1890/12-0826.1. PMID:24640535.
- Stefánsson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES J. Mar. Sci.* **53**(3): 577–588. doi:10.1006/jmsc.1996.0079.
- Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E., and Semmens, B.X. 2019. The utility of spatial model-based estimators of unobserved bycatch. *ICES J. Mar. Sci.* **76**: 255–267. doi:10.1093/icesjms/fsy153.
- Sumaila, U.R., Cheung, W.W.L., Lam, V.W.Y., Pauly, D., and Herrick, S. 2011. Climate change impacts on the biophysics and economics of world fisheries. *Nat. Clim. Change*, **1**: 449–456. doi:10.1038/nclimate1301.
- Thorson, J.T. 2018. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. *Can. J. Fish. Aquat. Sci.* **75**(9): 1369–1382. doi:10.1139/cjfas-2017-0266.
- 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 J. Mar. Sci.* **74**: 1311–1321. doi:10.1093/icesjms/fsw193.
- Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J., and Kristensen, K. 2015a. Spatial factor analysis: a new tool for estimating joint species distributions and correlations in species range. *Methods Ecol. Evol.* **6**: 627–637. doi:10.1111/2041-210X.12359.
- Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015b. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES J. Mar. Sci.* **72**: 1297–1310. doi:10.1093/icesjms/fsu243.
- Thorson, J.T., Skaug, H.J., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J.H., and Benante, J.A. 2015c. The importance of spatial models for estimating the strength of density dependence. *Ecology*, **96**: 1202–1212. doi:10.1890/14-0739.1. PMID:26236835.
- Thorson, J.T., Fonner, R., Haltuch, M.A., Ono, K., and Winker, H. 2017. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. *Can. J. Fish. Aquat. Sci.* **74**: 1794–1807. doi:10.1139/cjfas-2015-0598.
- Torgo, L. 2010. Data mining with R: Learning with case studies. Chapman and Hall/CRC, Boca Raton, Fla.
- Venables, W.N., and Ripley, B.D. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fish. Res.* **70**(2): 319–337. doi:10.1016/j.fishres.2004.08.011.
- Wager, S., and Athey, S. 2017. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **113**: 1228–1242. doi:10.1080/01621459.2017.1319839.
- Wager, S., Hastie, T., and Efron, B. 2014. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. *J. Mach. Learn. Res.* **15**: 1625–1651.
- Ward, E.J., Holmes, E.E., Thorson, J.T., and Collen, B. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. *Oikos*, **123**: 652–661. doi:10.1111/j.1600-0706.2014.00916.x.
- Ward, E.J., Jannot, J.E., Lee, Y.W., Ono, K., Shelton, A.O., and Thorson, J.T. 2015. Using spatiotemporal species distribution models to identify temporally evolving hotspots of species co-occurrence. *Ecol. Appl.* **25**: 2198–2209. doi:10.1890/15-0051.1. PMID:26910949.
- Watson, J.T., Essington, T.E., Lennert-Cody, C.E., and Hall, M.A. 2009. Trade-offs in the design of fishery closures: Management of silky shark bycatch in the Eastern Pacific Ocean Tuna fishery. *Conserv. Biol.* **23**: 626–635. doi:10.1111/j.1523-1739.2008.01121.x. PMID:19040650.
- Welling, S.H., Refsgaard, H.H.F., Brockhoff, P.B., and Clemmensen, L.H. 2016. Forest floor visualizations of random forests [online]. ArXiv e-prints. <http://arxiv.org/abs/1605.09196>.
- Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **73**(1): 3–36.
- Wood, S.N. 2017. Generalized additive models: An introduction with R (2nd ed.). Chapman & Hall/CRC, Boca Raton, Fla.

Appendix A

Appendix Figures A1–A6 appear on the following pages.

Fig. A1. Binomial component predictive performance (AUC) for the three random forest (RF) submodels for all six species. Modifications designed for data with severely imbalanced classes, down sampling and SMOTE, outperformed the default algorithm for the four species with high or low bycatch rates (BLUE, LOGG, LEATH, and YEYE). Species are sorted on the x axis by class imbalance (i.e., PHLB has the most moderate bycatch rate (29%), and YEYE has the most extreme bycatch rate (0.3%). For blue sharks, the minority class is negative (sets with zero bycatch, 3.8%). Species abbreviations: DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye rockfish, LOGG = loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark.

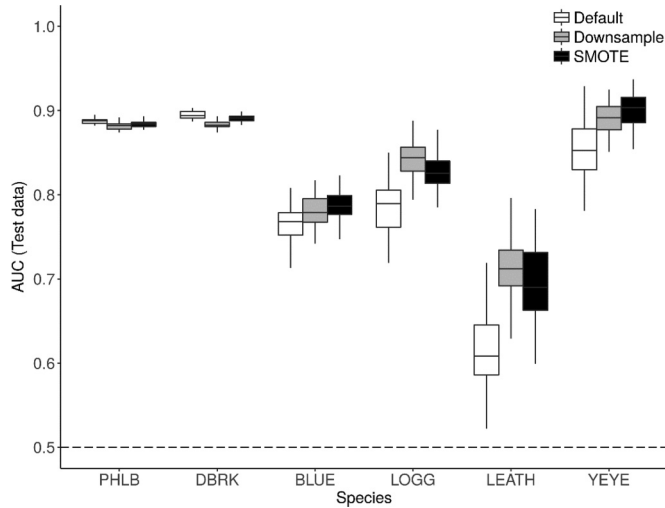
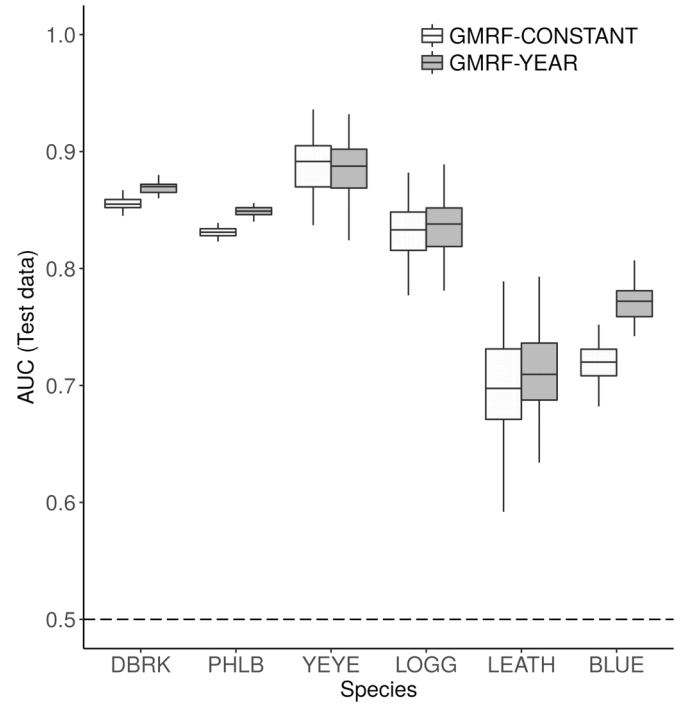


Fig. A2. Binomial component predictive performance (AUC) for the two GMRF models: CONSTANT (white, one random field constant across years) and YEAR (grey, random field fit for each year). Fitting random fields for each year (GMRF-YEAR) resulted in better model performance (higher AUC) for species with moderate to high bycatch rates (DBRK, PHLB, BLUE), but nearly identical performance for species with few bycatch occurrences (YEYE, LOGG, LEATH). Species abbreviations: DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye rockfish, LOGG = loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark.



Can. J. Fish. Aquat. Sci. Downloaded from cdsciencepub.com by NOAA CENTRAL on 06/05/23 For personal use only.

Fig. A3. GMRF-CONSTANT random field for bycatch probability of the three US West Coast groundfish species (DBRK = darkblotched rockfish, PHLB = Pacific halibut, and YEYE = yelloweye rockfish). The GMRF-CONSTANT model estimates one spatial field constant across years from 2003 to 2012. Fitting a complex spatial field is less useful for species with very low bycatch rates (e.g., yelloweye rockfish, right panel, nonzero bycatch in 143 out of 42 787 tows). In these cases, simply including linear and quadratic effects of latitude and longitude could achieve nearly the same result. [Colour online.]

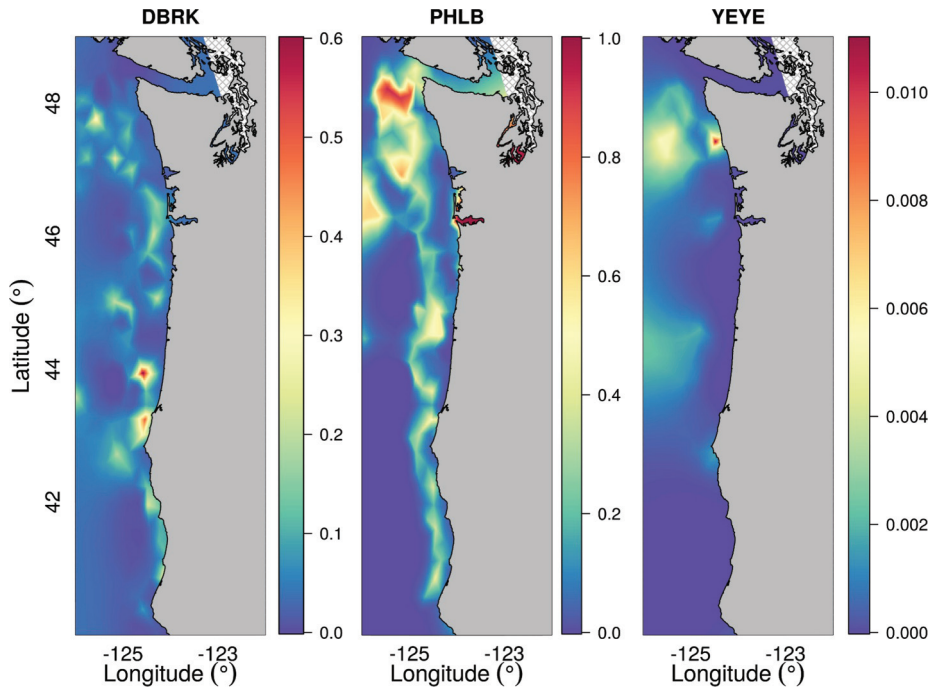


Fig. A4. Bycatch-to-target species catch ratio achieved for each species by using the binomial component of each delta-model to predict and remove fishing sets in the test data, relative to the bycatch-to-target ratio with no fishing sets removed. Lines show median of 50 cross-validation runs for each model class (fivefold cross-validation repeated 10 times). Lines are noisier for species with few bycatch events (e.g., YEYE, LOGG, and LEATH). Species abbreviations: DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye rockfish, LOGG = loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark. [Colour online.]

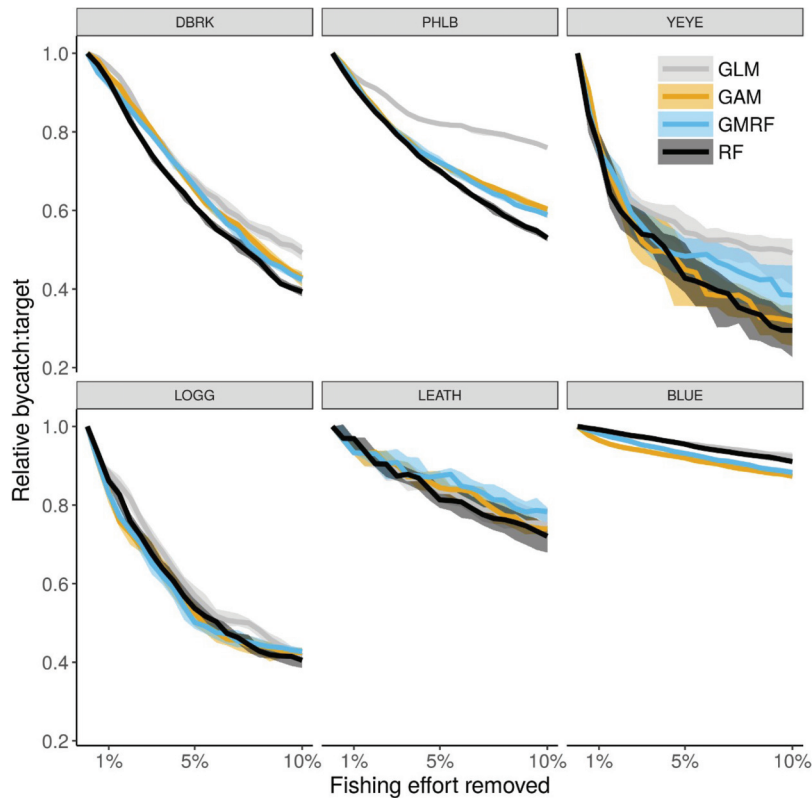


Fig. A5. Distribution of blue shark bycatch by latitude. Above 41.5°N (red line), there were few observed fishing sets (26 out of 16 total), and these resulted in higher and more variable bycatch. [Colour online.]

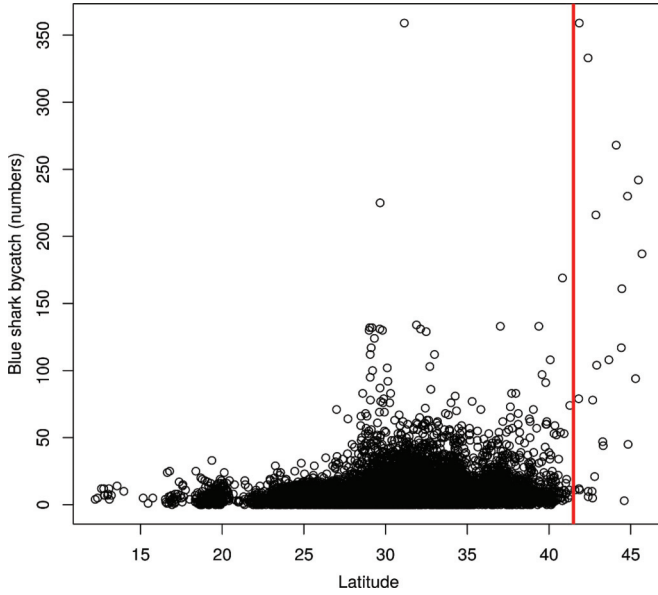
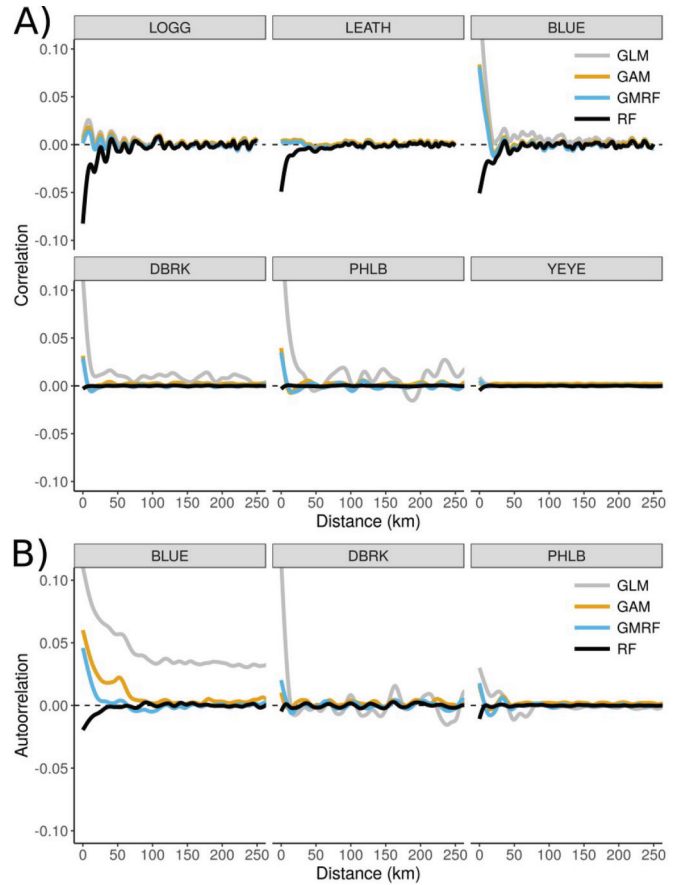


Fig. A6. Spatial spline correlograms of residuals from the (A) binomial and (B) positive components of the delta-models. Positive and negative values indicate positive and negative spatial autocorrelation (Moran's I), respectively, and $y = 0$ is the expected value under the null hypothesis of no spatial autocorrelation. The distance where the autocorrelation intersects ($y = 0$) is the decorrelation distance, which is roughly 40 km for the Hawaii longline species (blue shark = BLUE, loggerhead turtle = LOGG, and leatherback turtle = LEATH) and 5 km for the West Coast groundfish species (darkblotched rockfish = DBRK, Pacific halibut = PHLB, yelloweye rockfish = YEYE). Calculations were made by package "ncf" version 1.2-5 (Bjornstad and Falck 2001). [Colour online.]



Can. J. Fish. Aquat. Sci. Downloaded from cdsciencepub.com by NOAA CENTRAL on 06/05/23 For personal use only.