1 **Title:**

2 Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution

3 model frameworks

4

5 **Authors and affiliations:**

6 Brian C. Stock[1*], Eric J. Ward[2], Tomoharu Eguchi[3], Jason E. Jannot[4], James T. Thorson[4], Blake

7 E. Feist[2], Brice X. Semmens[1]

8

9 *Corresponding: Brian Stock, b1stock@ucsd.edu, +1 425-919-7879

10   Scripps Institution of Oceanography, UCSD

11   9500 Gilman Drive, mail code 0202

12   La Jolla, CA 92093

13 [1]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California

14   92093 USA, b1stock@ucsd.edu, bsemmens@ucsd.edu

15 [2]Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries

16   Service, National Oceanographic and Atmospheric Administration, 2725 Montlake Blvd E,

17   Seattle, WA 98112 USA, eric.ward@noaa.gov, blake.feist@noaa.gov

18 [3]Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and

19   Atmospheric Administration, 8901 La Jolla Shores Drive, La Jolla, CA 92037 USA,

20   tomo.eguchi@noaa.gov

21 [4]Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center,

22   National Marine Fisheries Service, NOAA, 2725 Montlake Boulevard East, Seattle, WA

23   98112 USA, jason.jannot@noaa.gov, james.thorson@noaa.gov

24

25 **Abstract**

26 Spatiotemporal predictions of bycatch (i.e. catch of non-targeted species) have shown promise as

27 dynamic ocean management tools for reducing bycatch. However, which spatiotemporal model

28 framework to use for generating these predictions is unclear. We evaluated a relatively new

29 method, Gaussian Markov random fields (GMRFs), with two other frameworks, generalized

30 additive models (GAMs) and random forests. We fit geostatistical delta-models to fisheries

31 observer bycatch data for six species with a broad range of movement patterns (e.g. highly

32 migratory sea turtles vs. sedentary rockfish) and bycatch rates (percent of observations with non-

33 zero catch, 0.3-96.2%). Random forests had better interpolation performance than the GMRF and

34 GAM models for all six species, but random forests performance was more sensitive when

35 predicting data at the edge of the fishery (i.e. spatial extrapolation). Using random forests to

36 identify and remove the 5% highest bycatch risk fishing events reduced the bycatch-to-target

37 species catch ratio by 34% on average. All models considerably reduced the bycatch-to-target

38 ratio, demonstrating the clear potential of species distribution models to support spatial fishery

39 management.

40

45

Introduction

Bycatch—catch of non-targeted 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 bycatch in the Alaska pollock 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 reasons, the fishing community is interested in tools to reduce bycatch.

One such tool 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 (DOM), 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; Eguchi et al. 2017; Hazen et al. 2016; Howell et al. 2008, 2015; Lewison et al. 2015). It is not 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 datasets

69 offer a wide range of occurrence rates and densities. Thus, in addition to providing guidance for

70 spatial bycatch management, large bycatch datasets are excellent testbeds for evaluating SDM

71 performance more generally.

72 *Species distribution models (SDMs)*

73   SDMs can be coarsely divided into parametric, semiparametric, and nonparametric

74 approaches. Generalized linear models (GLMs) are one of the simplest parametric approaches

75 used to understand species distributions and their relationships with biotic and abiotic covariates

76 (Venables and Dichmont 2004). GLMs predict the response variable, Y_i (species

77 presence/absence or abundance at location i), by specifying a probability distribution and link

78 function:

79 $$Y_i \sim \text{distribution with mean } \mu_i, \qquad g(\mu_i) = \eta_i,$$

80 with linear predictor

81 $$\eta_i = \mathbf{X_i}\boldsymbol{\beta}, \tag{1}$$

82 where $\mathbf{X_i}$ is a vector of covariate values for location i, and $\boldsymbol{\beta}$ is a vector of coefficients to be

83 estimated. GLMs can permit nonlinear relationships between the covariates and response by

84 including transformations of the covariates, e.g. polynomial terms $\eta_i = \beta_0 + \beta_1\mathbf{X_{1i}} + \beta_2\mathbf{X_{1i}^2} +$

85 $\beta_3\mathbf{X_{1i}^3} + \cdots$, or by discretizing continuous covariates and treating them as categorical variables.

86   Generalized additive models (GAMs) extend the GLM framework by allowing the linear

87 predictor to include smooth functions of the covariates (Guisan and Thuiller 2005; Wood 2017).

88 GAMs are often referred to as semiparametric, since the smoothers do not have a specified

89 functional form but do have associated parameters that are estimated using penalized likelihood

90 (Wood 2011; Guélat and Kéry 2018). The ability of GAMs to incorporate complex, non-linear

91 covariate effects, as well as improvements to computing power and software, has led to their

92   wide adoption in fisheries and ecology in the last decade (Becker et al. 2014; Leathwick et al.

93   2006; Li and Pan 2011; Watson et al. 2009). Extending the linear predictor in Equation 1 to

94   include a 2-dimensional spline, $f()$, on the geographical coordinates of location i, $\mathbf{s_i}$, specifies a

95   GAM:

$$\eta_i = \mathbf{X_i\beta} + f(\mathbf{s_i}). \tag{2}$$

97   Equation 2 is estimated by penalized likelihood maximization, which balances smoothness and

98   fit to the data by penalizing the curvature (i.e. integral of the squared second derivative) of $f(\mathbf{s_i})$

99   (Wood 2017). Kammann and Wand (2003) refer to Equation 2 as a 'geoadditive' model, and have

100 shown that this is mathematically equivalent to explicitly modeling spatial correlation with

101 random effects, $\mathbf{u}$:

$$\eta_i = (\mathbf{X\beta} + \mathbf{Zu})_i \tag{3}$$

103 (Diggle et al. 1998; Kneib et al. 2008; Péron et al. 2011; Fahrmeir et al. 2013; Guélat and Kéry

104 2018). When the spatial random effects are assumed to follow a zero-mean multivariate normal

105 distribution, Equation 3 can be written as:

$$\begin{aligned} \eta_i &= \mathbf{X_i\beta} + \boldsymbol{\varepsilon}(\mathbf{s_i}), \\ \boldsymbol{\varepsilon}(\mathbf{s}) &\sim \mathrm{MVN}(0, \boldsymbol{\Sigma}), \end{aligned} \tag{4}$$

107 where $\boldsymbol{\varepsilon}$ is a Gaussian field (Kneib et al. 2008). Analogous to how the curvature of the spline is

108 penalized when estimating the GAM, the correlation function that defines $\boldsymbol{\Sigma}$ acts as a penalized

109 spatial smoother in the Gaussian field model—nearby locations are more highly correlated, and

110 thus more smoothed, than distant locations (Fahrmeir et al. 2013). Gaussian fields are attractive

111 because they directly model spatial correlation, but applications have historically been limited to

112 smaller datasets because inverting the covariance matrix, $\boldsymbol{\Sigma}$, makes them computationally intense

113 (Lindgren et al. 2011). In summary, both the GAM and Gaussian field model account for spatial

114 autocorrelation not explained by environmental covariates, are semiparametric, mathematically

115 equivalent, and typically fit using penalized likelihood that optimizes spatial smoothing.

116 Gaussian fields are defined in continuous space but can be approximated by discrete

117 Gaussian Markov random fields (GMRFs, Lindgren et al. 2011). GMRFs have increasingly been

118 used to model species distributions as advances in computing power and software

119 implementation have allowed ecologists to apply them to large datasets. Among other

120 advantages, the GMRF approach can be implemented by integrated nested Laplace

121 approximation, which is faster than other methods of Bayesian inference (i.e. Markov chain

122 Monte Carlo) and allow GMRF approximations of Gaussian fields to be computationally feasible

123 (Rue et al. 2009). GMRF models have shown promise in assessing relationships between habitat

124 and distribution (Illian et al. 2013), the effects of interspecific relationships such as density

125 dependence (Thorson et al. 2015c), as well as the relationships between multiple co-occurring

126 taxa (Ward et al. 2015). From a quantitative standpoint, GMRF models have been shown to

127 estimate population abundance trends with greater precision and accuracy compared to non-

128 spatial models (Thorson et al. 2015b).

129 Clearly, GMRFs are intimately related to GAMs, since GMRFs approximate the

130 Gaussian field model (Eqn. 4), which is an alternative parameterization of the GAM in Equation

131 2. The spatial smoothing terms for both GAMs and GMRFs can be defined on a sphere or differ

132 according to spatial direction (anisotropy can be included in GAMs by using tensor product

133 smooths, and Σ can be both non-stationary and anisotropic in GMRFs). GAMs and GMRFs

134 differ, however, in two regards. First, the spatial smoothing term appears as a mean trend for

135 GAMs and as a covariance matrix for GMRFs. This distinction may not be important for

136 modelling species distributions, and researchers may prefer the method that reflects their view of

137 how spatial autocorrelation arises in the problem at hand. For instance, GMRFs would be the

138   more natural framework if the spatial variation remaining after including environmental

139   covariates is considered random. Second, and more importantly, the different parameterizations

140   lead to different estimation methods and software implementations. Since spatial models for

141   large datasets contain many parameters, numerically efficient implementations are crucial

142   (Fahrmeir et al. 2013). Several R packages fit GAMs, the most popular being 'mgcv,' which uses

143   generalized cross-validation to estimate smooth terms by default (Wood 2017). GMRFs can be

144   fit via 'INLA,' which uses a Bayesian framework and estimates models by integrated nested

145   Laplace approximation (Lindgren and Rue 2015). In theory, 'mgcv' and 'INLA' should be quite

146   similar. In practice, however, differences in approximation methods, runtime, convergence

147   criteria, ease of use, and default settings may impact model predictions.

148   Most nonparametric approaches ecologists use to model species distributions have

149   evolved from machine learning algorithms (Hastie et al. 2009; Olden et al. 2008). These data-

150   driven approaches include random forests (RF, Breiman 2001; Cutler et al. 2007), MaxEnt

151   (Phillips et al. 2006; Phillips and Dudík 2008), and support vector machines (Drake et al. 2006).

152   In this analysis we highlight RF because 1) data in our application—fisheries bycatch—contain

153   true absences, whereas MaxEnt is designed for presence-only data, and 2) RF is widely used and

154   has shown good predictive performance in SDM testing (Prasad et al. 2006; Marmion et al.

155   2009; Scales et al. 2016). The RF algorithm predicts the response by constructing m regression

156   (or classification) trees and averaging their predictions (Breiman 2001). Each individual tree

157   begins with all observations and then iteratively partitions the data by splitting along one

158   covariate (e.g. depth > 100 m versus depth ≤ 100 m), choosing the covariate and split point that

159   minimizes the sums of squares error at each node (where the predicted response at each node is

160   the mean of observations within the node, Breiman et al. 1984). The process continues until each

161   terminal node contains less than a specified number of observations. Individual trees are simple

162   and computationally cheap but are also unstable (i.e. sensitive to slight alterations in the data)

163   and sub-optimal at prediction (i.e. they are "weak learners"), because they only allow rectangular

164   partitions of covariate space. The RF algorithm increases predictive performance by reducing the

165   correlation between trees, which is accomplished via two processes: 1) fitting each tree to a

166   bootstrap sample of the original data, and 2) at each split, randomly selecting a subset of

167   covariates to consider (Kuhn and Johnson 2013). This works because reducing the correlation

168   between individual trees reduces the correlation of their errors, which therefore reduces the

169   predictive error of their average, the RF estimate.

170   Random forests are popular because they are simple to use (few parameters to tune and

171   the default values work well in most cases), robust to the inclusion of many non-informative

172   covariates generate accurate predictions, designed to not overfit, and seamlessly accommodate

173   missing data (Biau and Scornet 2016). Compared to parametric and semiparametric models, RF

174   will often have better out-of-sample (i.e. cross-validated) prediction performance due to their

175   ability to estimate more complex patterns, as non-linearity and interactions are inherent in their

176   construction (Elith and Leathwick 2009). However, this data-driven complexity does come at the

177   cost of model interpretability, and this is one of the main factors limiting the adoption of RF—

178   and machine learning methods more generally—by ecologists (Olden et al. 2008). Three other

179   disadvantages of RF are the difficulty of generating uncertainty estimates with well-understood

180   properties, analyzing model diagnostics, and specifying constraints on model fit (e.g., we may

181   wish yearly estimates to be independent, which can be specified in parametric models).

182   *Study objectives*

183   The primary objective of this paper is to compare the performance of GAMs, GMRFs

184   and RF in a predictive framework using cross-validation (Kuhn and Johnson 2013; Roberts et al.

185   2017). There has been an increased emphasis in ecology on evaluating and selecting models

186   based on their ability to predict out-of-sample data (Hooten and Hobbs 2015), and one of the

187   advantages of this approach is that nonparametric and parametric models can be compared (Ward

188   et al. 2014). While each of these model frameworks have individually been applied to understand

189   spatiotemporal trends in fisheries bycatch (GAMs: Becker et al. 2014; Hazen et al. 2016;

190   McCracken 2004; Watson et al. 2009; GMRFs: Breivik et al. 2016, 2017; Cosandey-Godin et al.

191   2015; RFs: Carretta et al. 2017; Eguchi et al. 2017; Pons et al. 2009), their predictive

192   performance has not been tested in a comparative study.

193   Our next objective is to evaluate the utility of using SDM predictions of bycatch risk as a

194   tool to reduce bycatch in fisheries. Beyond abstract performance metrics, we compare the

195   models' capabilities to reduce the bycatch-to-target species catch ratio, create spatial bycatch risk

196   maps, and estimate effects of covariates.

197   The final objective of our analysis is to evaluate model transferability, the ability to

198   extrapolate, or predict beyond the range of observed data. Traditional cross-validation only

199   measures a model's ability to interpolate, i.e. estimate values within the range of observations,

200   because it randomly chooses data to withhold for testing. SDMs that are more data-driven and

201   complex have been shown to have better interpolation performance but be worse at spatial

202   extrapolation (Araújo and Rahbek 2006; Heikkinen et al. 2012; Randin et al. 2006). In other

203   words, one model may have higher predictive performance in the core fishing area with abundant

204   data, yet underperform other models in areas with sparse sampling coverage. Since we wish to

205   evaluate using SDM predictions of bycatch risk as a spatial management tool, it is important to

206   assess how sensitive the predictions are to spatial location. Predictions in areas with few data are

207 more sensitive to model misspecification and overfitting, and therefore caution is especially

208 warranted for complex, nonparametric approaches such as RF (Merow et al. 2014).

209 **Methods**

210 *Fisheries observer data*

211 Collecting reliable bycatch data depends on fisheries observer programs, where on-board

212 observers enumerate and record the species caught (as well as fishing location, gear type, time,

213 and other relevant information). To explore the performance of species distribution models

214 across taxa, we used two datasets from United States fisheries observer programs in the Pacific

215 Ocean with high observer coverage. The first dataset was from the West Coast Groundfish

216 Observer Program (WCGOP) at the Northwest Fisheries Science Center (NWFSC, Bellman et al.

217 2010). The WCGOP dataset contained records of 42 786 commercial bottom trawls from 2003-

218 2012 off the west coast of the USA, primarily targeting groundfish such as Dover sole

219 (*Microstomus pacificus*), thornyheads (*Sebastolobus* spp.), sablefish (*Anoplopoma fimbria*), and

220 rockfish (*Sebastes* spp., Fig. 1a). Observers recorded haul duration, location, date, time, depth,

221 gear type, and catch (which includes at-sea discarded bycatch; for details see NWFSC 2016).

222 Observer coverage was approximately 20% from 2003-2010 under limited access management,

223 with 100% coverage starting in 2011 with the transition to an individual fishing quota (IFQ)

224 system. In the pre-IFQ era, fishermen were not permitted to land rebuilding species (i.e.

225 populations declared overfished with management plans to rebuild to sustainable levels), so we

226 defined bycatch as only at-sea discards. Under the IFQ system fishermen can land a low quota of

227 rebuilding species, so we considered bycatch to be the sum of discarded and retained catch for

228 non-target species.

229   The second dataset was from the Hawaii longline (HILL) fishery, monitored by the

230   Pacific Islands Regional Observer Program (PIROP 2014), which has recorded fishing location,

231   date, time, sea surface temperature (SST), gear characteristics, and catch of longline sets from

232   1994-2014. The Hawaii longline fleet is divided into two sectors, one targeting tuna (*Thunnus*

233   *spp.*) and the other swordfish (*Xiphias gladius*), with distinct gear configurations and

234   spatiotemporal effort patterns, both of which affect interaction rates with bycatch species (Li and

235   Pan 2011). We modeled 16 714 observations from the shallow-set swordfish fishery in 1994-

236   2001 and 2005-2014 (Fig. 1b), distinguishing between sets targeting swordfish and tuna by the

237   number of hooks between surface floats (following Li and Pan 2011). Concerns over bycatch of

238   protected species, particularly of loggerhead (*Caretta caretta*) and leatherback sea turtles

239   (*Dermochelys coriacea*), motivated the closure of the swordfish fishery from 2001 to 2004. This

240   led to two important differences between the data from 1994-2001 and 2005-2014. First, sea

241   turtle bycatch rates have been an order of magnitude lower in the later period, the result of

242   stricter regulations and modifying hooks (J to circle hooks) and bait types (squid to fish; Gilman

243   et al. 2007). Second, observer coverage increased from roughly 5% to 100% (Howell et al.

244   2008).

245   Model performance may be linked to species' movement patterns, because species that

246   move less (or whose movement patterns do not change in time) may not need a spatiotemporal

247   model. Instead, a time-constant spatial model may be adequate. To ascertain whether differences

248   in SDM performance were related to movement pattern or bycatch rate (i.e. % observations with

249   non-zero catch), we selected three bycatch species from each dataset: blue shark (*Prionace*

250   *glauca*), loggerhead sea turtle, and leatherback sea turtle from the Hawaii longline fishery, and

251   Pacific halibut (*Hippoglossus stenolepis*), darkblotched rockfish (*Sebastes crameri*), and

252   yelloweye rockfish (*Sebastes ruberrimus*) from the West Coast groundfish trawl fishery. These

253 species widely differ in their bycatch rates (96.2%, 0.7%, 0.3%, 28.9%, 17.9%, and 1.4%,

254 respectively), habitat preferences, and movement patterns. For instance, rockfish are relatively

255 sedentary and closely associated with rocky bottom habitat, whereas halibut exhibit seasonal and

256 long-distance migrations (Skud 1977; Gunderson 1997). In contrast to the groundfish, blue

257 sharks and sea turtles inhabit the open ocean and range much more widely (Benson et al. 2011;

258 Kobayashi et al. 2008; Nichols et al. 2000).

259   While both datasets include periods with 100% observer coverage, they also span periods

260 with partial coverage. This is relevant since the models assume that the data represent a random

261 sample of the studied fishery, i.e. each fishing event has an equal probability of being observed.

262 For several reasons, it is difficult for observer programs to achieve random sampling: a list of

263 trips and their departures often does not exist far in advance, certain vessels may not be able to

264 accommodate observers, observers may not always be available, and fisher behavior can change

265 when observers are on board (Hall 1999; Liggins et al. 1997; McCracken 2004). The WCGOP

266 data from years with 20% coverage are likely to be representative of the fishery, because the

267 WCGOP stratified sampling by port group, vessel, and 2-month blocks with the goals of

268 sampling all vessels for two months in each year and discouraging changes to fishing behavior

269 when observers were on-board (NWFSC 2006). It is less likely that this was true for the 1994-

270 2001 HILL data. Nevertheless, we included data from periods with partial coverage because

271 there were very few observations of non-zero catch for rarely encountered species in the years

272 with full coverage (yelloweye rockfish: 38, loggerhead turtle: 89, leatherback turtle: 82), and in

273 many cases, bycatch of these 'rare-event' species are often of highest management concern

274 (Martin et al. 2015).

275 *Environmental covariates*

276　In addition to the locations of observed fishing, we considered several covariates that may help

277　explain the likelihood of bycatch events. For the WCGOP dataset, we included fishing depth, day

278　of year, sea surface temperature (SST) anomaly, distance to rocky habitat, size of nearest rocky

279　patch, predicted occurrence from survey data, and whether the trawl occurred in or near a

280　Rockfish Conservation Area (RCA). RCAs are large areas along the U.S. West Coast closed to

281　fishing designed primarily to reduce bycatch of overfished rockfish, such as two of the species

282　we considered. RCA boundaries have changed by and within years, and are defined by latitude,

283　date, and depth (NOAA Fisheries West Coast Region 2015). Trawls were determined to be inside

284　or outside of an RCA based on the trawl date, average position of trawl start and end, and bottom

285　depth (calculated via bathymetry from NOAA National Centers for Environmental Information

286　2015). We included linear and quadratic terms for fishing depth and SST anomalies following

287　Shelton et al. (2014). Depth was recorded by on-board observers, while SST anomalies were

288　measured via satellite. For each trawl, we collected daily SST anomalies on a 0.25º grid and used

289　bilinear interpolation to create SST anomalies corresponding to each trawl location

290　(http://www.esrl.noaa.gov/psd/, Reynolds et al. 2007). Rocky habitat data were from NMFS

291　(2013), calculated as per Shelton et al. (2014). Finally, we used the above covariates to fit a

292　geostatistical binomial GLMM to fisheries-independent trawl survey data (Bradburn et al. 2011,

293　modeled as in Shelton et al. 2014), and applied this model to predict bycatch occurrence at the

294　fishing times and locations in the observer dataset. These survey-predicted occurrence

295　probabilities were included as another linear covariate. All environmental covariates were

296　centered before model estimation.

297　　　　The only available environmental covariate for the HILL dataset was observer-recorded

298　SST, and therefore we fit the HILL models with covariates of standardized SST, SST^2, and day of

299   year.

300   *Statistical models*

301   As is common for species distribution data, five of the six species exhibited large proportions of

302   zero catches. We followed the hurdle- or delta-model approach to this complication, which is

303   commonly applied in ecology and fisheries (Pennington 1983; Maunder and Punt 2004). Delta-

304   models separate the observed catches, Y_i, into two processes: a 'binomial' component for the

305   probability of non-zero catch, π_i, and a 'positive' component for the mean catch density given the

306   catch is non-zero, μ_i:

307
$$Z_i \sim \text{Bernoulli}(\pi_i)$$
$$Y_i \sim Z_i\, h(\mu_i)$$
(5)

308   where Z_i is a binary variable that equals 1 if the species was caught and 0 if it was not, and $h()$ is

309   a distribution to be specified (e.g. lognormal, gamma). Splitting the modeling into these two

310   components can be advantageous because different mechanisms may affect one component but

311   not the other (e.g. a habitat quality covariate may be a significant predictor of catch rate, but not

312   occurrence).

313   We applied a total of eight delta-models with varying spatial structure to each of the six

314   species included in our analysis (Table 1). Bycatch of yelloweye rockfish, loggerhead turtles, and

315   leatherback turtles were extremely rare events (0.3-1.4% non-zero observations) with too few

316   multiple-individual catches to meaningfully fit the positive component. All analyses were

317   conducted using R v3.4.1 (R Core Team 2017), with the following libraries: 'mgcv' was used to

318   implement GLMs and GAMs (v1.8-17, Wood 2017); 'randomForest' (v4.6-12, Liaw and Wiener

319   2002), 'DMwR' (v0.4.1, Torgo 2010), and 'forestFloor' (v1.9.5, Welling et al. 2016) were used to

320   fit RFs; and 'INLA' was used to fit the GMRF models (v0.0-1485844051, Lindgren and Rue

321   2015). We assessed model fit with plots of covariate-response relationships, predicted versus

322   observed response in out-of-sample data, spatial residual maps, and spatial correlograms

323   (Moran's I, package 'ncf' v1.2-5, Bjørnstad and Falck 2001). Code to fit each of the models is

324   provided at https://github.com/brianstock/spatial-bycatch.

325   Our first model was a delta-GLM with linear and quadratic effects of the environmental

326   covariates (which are intrinsically spatially correlated), but without any spatial terms—neither

327   geographic coordinates nor spatial autocorrelation for residual errors. As in Guélat and Kéry

328   (2018), the delta-GLM served as a baseline that allowed us to evaluate the value of adding

329   spatial terms in the subsequent models, which were fit using the same covariates and only differ

330   in how they include spatial information. The delta-GLM fits the observed bycatch in fishing

331   event i, Y_i, as in Eqns. 1 and 5, with binomial component determining the probability of non-zero

332   bycatch, π_i:

$$
\begin{aligned}
Z_i &\sim \text{Bernoulli}(\pi_i), \\
\text{logit}(\pi_i) &= \mathbf{X_i}\boldsymbol{\alpha},
\end{aligned}
\tag{6}
$$

334   and positive component for the mean catch density given the catch is non-zero, μ_i:

$$
\begin{aligned}
Y_i &\sim Z_i\, \text{Gamma}(\mu_i, k), \\
\log(\mu_i) &= \mathbf{X_i}\boldsymbol{\beta}.
\end{aligned}
\tag{7}
$$

336   where $\mathbf{X_i}$ is a vector of covariate values for location i, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of coefficients to be

337   estimated, and k is the shape parameter of the gamma distribution. The gamma distribution is

338   appropriate for positive, right-skewed data, and therefore is commonly used in the positive

339   component of delta-models for fisheries catch (Lecomte et al. 2013; Stefánsson 1996). While we

340   would not expect the GLM to outperform the models with explicit spatial terms, it is possible

341   that the (spatially-structured) environmental covariates could explain most of spatial structure in

342   the response. In that case, including spatial terms in the model (i.e. a 2-d spline as in Eqn. 2 or

343   covariance matrix as in Eqn. 4) would be unnecessary.

344   We fit two delta-GAM models that extend Eqns. 6 and 7 by adding a 2-dimensional

345   spline, $f()$, on the geographical coordinates of location i, $\mathbf{s_i}$, to both the binomial and positive

346   components, as in Eqn. 2:

$$
\begin{aligned}
Z_i &\sim \text{Bernoulli}(\pi_i), \\
\text{logit}(\pi_i) &= \mathbf{X_i}\boldsymbol{\alpha} + f_Z(\mathbf{s_i}), \\
Y_i &\sim \text{Gamma}(\mu_i, k), \\
\log(\mu_i) &= \mathbf{X_i}\boldsymbol{\beta} + f_Y(\mathbf{s_i}).
\end{aligned}
\tag{8}
$$

347

348   The first, "GAM-CONSTANT," includes one 2-d spline constant across years, with an offset

349   (fixed effect) for each year. This allows the mean bycatch probability and density to vary

350   temporally and spatially, but in the same pattern each year. The second, "GAM-YEAR," fits an

351   independent 2-d spline for each year, which allows the spatial pattern to vary between years

352   (Table 1).

353   As for the GAMs, we fit two delta-GMRF models which extend Eqns. 6 and 7 by

354   estimating the covariance between observed locations, s_i, as in Eqn. 4:

$$
\begin{aligned}
Z_i &\sim \text{Bernoulli}(\pi_i), \\
\text{logit}(\pi_i) &= \mathbf{X_i}\boldsymbol{\alpha} + \boldsymbol{\varepsilon_Z}(\mathbf{s_i}), \\
\boldsymbol{\varepsilon_Z}(\mathbf{s}) &\sim \text{MVN}\left(0, \mathbf{Q_Z^{-1}}\right), \\
Y_i &\sim \text{Gamma}(\mu_i, k), \\
\log(\mu_i) &= \mathbf{X_i}\boldsymbol{\beta} + \boldsymbol{\varepsilon_Y}(\mathbf{s_i}), \\
\boldsymbol{\varepsilon_Y}(\mathbf{s}) &\sim \text{MVN}\left(0, \mathbf{Q_Y^{-1}}\right),
\end{aligned}
\tag{9}
$$

355

356   where both $\mathbf{Q_Z^{-1}}$ and $\mathbf{Q_Y^{-1}}$ are defined to approximate stationary, isotropic Matérn covariances,

357
$$
\text{Cov}(\mathbf{s_1}, \mathbf{s_2}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\mathbf{s_1} - \mathbf{s_2}\|)^\nu K_\nu(\kappa\|\mathbf{s_1} - \mathbf{s_2}\|),
$$

358   K_ν is the modified Bessel function of the second kind and order $\nu > 0$, κ is the spatial scale

359   parameter, and $\boldsymbol{\varepsilon_Z}()$ and $\boldsymbol{\varepsilon_Y}()$ represent the estimated spatial fields using random effects. We used

360   the default Matérn smoothness, $\nu = 1$, and priors on parameters as implemented in R-INLA

361   (Lindgren and Rue 2015). Analogous to the GAM-CONSTANT and GAM-YEAR models, we fit

362 a "GMRF-CONSTANT" model with one random field constant across all years, and a "GMRF-

363 YEAR" model with a random field estimated for each year (Table 1). As for GAM-CONSTANT,

364 the GMRF-CONSTANT model includes fixed effect terms for each year, which allow for an

365 increase or decrease in the mean bycatch probability and density for each year while assuming

366 the spatial pattern is constant across years. The GMRF-YEAR model uses the simplest

367 spatiotemporal option in R-INLA, 'exchangeable,' which refers to the spatiotemporal structure—

368 the random fields in all years are uniformly correlated (as opposed to an autoregressive

369 spatiotemporal structure where nearby years are more correlated than distant years).

370   To include spatiotemporal effects in RFs, we added year (treated as a factor), latitude, and

371 longitude as covariates. For the positive component of the delta-model, we fit only one RF

372 model: "RF-BASE," following the original RF algorithm as described by Breiman (2001) and

373 implemented in the 'randomForest' R package (Liaw and Wiener 2002). For the binomial

374 component, we also fit two modifications to the original RF algorithm designed to improve

375 performance on imbalanced class data (i.e. proportions of 0s and 1s very unequal), because

376 several species showed strong class imbalance (e.g., yelloweye rockfish had 99.7% tows with

377 zero catch and only 0.3% tows with non-zero catch). Training a RF on such severely imbalanced

378 class data tends to produce models that predict the majority class well but performs poorly on the

379 minority class (Kuhn and Johnson 2013). The first approach was to *down-sample* the majority

380 class observations (e.g. 0s, tows with zero catch for yelloweye rockfish) when training the RF

381 such that each tree used a stratified bootstrap sample of the data with equal numbers of majority

382 and minority (e.g. 1s, tows with non-zero catch for yelloweye rockfish) class observations. We

383 implemented down-sampling in the "RF-DOWN" model, setting the sample size equal to one-

384 sixth the number of minority class observations. The second approach, "RF-SMOTE" (Synthetic

385 Minority Over-sampling Technique, Chawla et al. 2002), combined down-sampling of the

386 majority class with over-sampling of the minority class. Simply over-sampling the minority class

387 with replacement overfits the RF to the specific observed minority data and typically does not

388 significantly improve prediction of the minority class. Instead, SMOTE creates synthetic

389 minority class samples by generating random linear combinations of nearby observed minority

390 samples, i.e. if X_1 and X_2 are nearest neighbors with $X_1 = \{SST_1, Lat_1, Lon_1\}$ and $X_2 = \{SST_2,$

391 $Lat_2, Lon_2\}$, draw p from Uniform(0,1) and create $X_{syn} = \{p(SST_1 - SST_2) + SST_2, p(Lat_1 - Lat_2)$

392 $+ Lat_2, p(Lon_1 - Lon_2) + Lon_2\}$.

393 *Performance metrics*

394 We performed 5-fold cross-validation repeated 10 times, which allowed us to evaluate

395 semiparametric and nonparametric SDMs' predictive error on new data (Shmueli 2010; Kuhn

396 and Johnson 2013). We blocked by year (rather than systematically excluding a given year) to

397 account for temporal structure and mimic predictive performance for cases where a random

398 subset of samples in a given year are not observed (Roberts et al. 2017). Thus, we generated 50

399 test/train splits, where each selects 20% of the data from each year to reserve for testing, and fits

400 the models on the remaining 80% training data. After fitting the models to training data, we used

401 the fitted models to predict bycatch probability (binomial component of the delta-model) and

402 density (positive component of the delta-model) at the test locations. This gave us predicted and

403 observed values to compare parametric and nonparametric model performance: area under the

404 receiver operating characteristic curve (AUC) scores for the binomial component, and root-

405 mean-square error (RMSE) for the positive component. AUC can be interpreted as the

406 expectation that a model ranks a uniformly drawn random positive (bycatch event) as more likely

407 than a uniformly drawn random negative (non-bycatch event), and higher values indicate better

408 performance (Hand 2009). We tested for significance across- and within-species using Tukey's

409 HSD test.

410 *Use as tools to reduce bycatch*

411 Since AUC and RMSE are abstract metrics of model performance, we calculated a more

412 tangible measure: how much each of the models could possibly reduce the bycatch-to-target

413 species catch ratio if it were used to identify and remove high bycatch risk fishing events. In

414 other words, what does an AUC performance gap of 0.03 mean in terms of bycatch reduction

415 (keeping catch of the target species constant)? For each species and model fit, we rank-ordered

416 the bycatch probabilities predicted by the binomial component, identified the X% of fishing

417 events with the highest bycatch probabilities (X% from 0-10%), removed both bycatch and target

418 catch from those events, and calculated the resultant change in bycatch-to-target species catch

419 ratio.

420 To be useful bycatch management tools, models also need to generate reasonable spatial

421 predictions, covariate effects, and pass diagnostic checks on spatial autocorrelation and residual

422 patterns. To investigate spatial predictions and their uncertainty, we calculated the posterior mean

423 and variance at each point on a prediction grid, using the R package 'mgcv' for GLMs and

424 GAMs (Wood 2017), and 'INLA' for GMRFs (Lindgren and Rue 2015). It is not possible to

425 calculate posterior distributions for RF, but we calculated standard errors using the infinitesimal

426 jackknife estimator ('randomForestCI', Wager et al. 2014). To compare covariate effects, we

427 plotted marginal posterior distributions for GLMs, GAMs, and GMRFs. To visualize RF

428 covariate effects, we used feature contributions calculated by the 'forestFloor' package (Welling

429 et al. 2016). Lastly, we investigated the models' abilities to reduce spatial autocorrelation using

430 spline correlograms ('ncf', Bjornstad and Falck 2001) and binned residual plots.

Transferability

432   In order to evaluate model transferability, i.e. spatial extrapolation performance, we

433   conducted a second cross-validation blocking on spatial data density. Whereas our first cross-

434   validation procedure partitioned the data into training and testing sets *randomly*, here we

435   constructed test/train splits by ordering fishing locations relative to data density. We calculated a

436   bivariate kernel density estimate at each of the observed fishing locations ('bkde2D' function in

437   'KernSmooth' R package) and sequentially used the lowest 0.5%, 1%, 2%, 5%, 10%, and 20%

438   density locations as test datasets. We then fit the models using only the core, data-rich area of

439   each fishery, and computed AUC and RMSE for model predictions at the test locations in

440   sparsely sampled areas.

441   **Results**

442   *Performance metrics*

443   Across the six species, RF provided better bycatch predictions than both GAMs and

444   GMRFs in the binomial (higher AUC, Fig. 2a) and positive (lower RMSE, Fig. 2b) components

445   of the delta-model. However, the magnitude of this performance advantage varied by species,

446   and in some cases the within-species differences were not significant. GMRFs outperformed

447   GAMs in the binomial component for the three species with moderate-high bycatch rates ($p <$

448   0.05, Tukey's HSD; darkblotched rockfish, Pacific halibut, and blue shark in Fig. 2a), but

449   differences in AUC were not significant for the three rarely caught species ($p > 0.05$, Tukey's

450   HSD; yelloweye rockfish, loggerhead turtle, and leatherback turtle in Fig. 2a). The variability in

451   model performance among cross-validation runs was similar within a given species but varied

452   greatly between species; variability in the binomial component was lowest for species with

453   moderate bycatch rates (darkblotched rockfish and Pacific halibut in Fig. 2a), and in the positive

454 component it was lowest for species with high bycatch rates (blue shark in Fig. 2b).

455   RF modifications designed for data with imbalanced classes, down-sampling and

456 SMOTE, outperformed the original RF algorithm for the four species with high or low bycatch

457 rates (Fig. A1). Of the semiparametric models, GMRF-YEAR that allowed for time-varying

458 spatial effects performed the best (Figs. 2 and A2), whereas the binomial GAM-YEAR that

459 estimated 2-d splines by year failed to converge. For rare bycatch species with few positive

460 occurrences, such as yelloweye rockfish, GMRF models that allowed for time-varying spatial

461 effects offered no improvement over the time-constant spatial models (Fig. A2). For these

462 species, the estimated GMRF spatial field was less complex than for species with higher bycatch

463 rates (Fig. A3).

464 *Use as tools to reduce bycatch*

465   When using the models to identify and remove high bycatch risk fishing events, RF also

466 performed best. Averaged across the six species, RF reduced the bycatch-to-target species catch

467 ratio by 8% when removing 1% of fishing effort, 34% when removing 5% of fishing effort, and

468 50% when removing 10% of fishing effort (Fig. 3). Bycatch predictability as measured by

469 bycatch-to-target ratio reduction generally agreed with the traditional performance metrics (AUC

470 and RMSE) and varied substantially among the six species (Fig. A4).

471   Bycatch risk maps produced by each of the models were similar in some respects. Taking

472 blue shark bycatch density an example, all models predicted higher and more variable bycatch in

473 the northwest area of the fishery, and lower bycatch between 25-30°N and 205-220°E (Fig. 4a-

474 d). Maps revealed artifacts of their construction: the mesh triangulation is evident in the GMRF

475 variance (Fig. 4g), and the sharp gradients in the RF mean (Fig. 4d) and variance (Fig. 4h) are a

476 consequence of RF trees splitting on latitude and longitude. As expected, the uncertainty of

477   GMRF predictions was lower in areas of high data density (compare Figs. 1b and 4g).

478   Uncertainty in RF predictions did not follow this trend—RF variance was extremely high above

479   41.5°N and moderately high between 30-33°N (Fig. 4h). However, this was consistent with the

480   data, as there were few observed sets north of 41.5°N and these had higher and more variable

481   bycatch (Fig. A5).

482   All models estimated similar covariate effects, as demonstrated for darkblotched rockfish

483   (Fig. 5). The main covariate effects were as expected: probability of bycatch increased with

484   higher survey-predicted occurrence, increased for tows inside or near RCAs, and showed an

485   optimal depth range of 100-250 fathoms.

486   For all species, residuals were typically more variable between cross-validation fits as

487   model complexity increased (RF > GMRF > GAM > GLM, Figs. S1-S18). Most models did not

488   exhibit spatial patterns in residuals, although residuals in the positive component of the delta-

489   model were larger and more variable for blue shark and Pacific halibut north of 40°N and 47°N,

490   respectively (Figs. S1-S2, S11-S12). GMRFs and GAMs had similarly lower residual spatial

491   autocorrelation compared to the baseline GLMs, with one exception where the GMRF reduced

492   spatial autocorrelation more than the GAM (blue shark positive model, Fig. A6). RF generally

493   had the lowest spatial autocorrelation, and it was negative instead of positive at short distances.

494   The Hawaii longline species had greater decorrelation distances than the West Coast groundfish

495   species (distance at which spatial autocorrelation goes to zero, 40 km vs. 5 km, Fig. A6).

496   *Transferability*

497   As expected, all models performed worse at spatial extrapolation compared to

498   interpolation, i.e. worse at predicting locations outside the core area of the fishery when trained

499   using observations in areas of highest data density (Fig. 6). GLMs and GAMs generally

500   performed worse in this test than GMRFs and RFs. GMRFs had lower RMSE than RFs in the

501   positive component of the delta-model but there was no difference in AUC in the binomial

502   component (Wilcoxon signed-rank test, $p = 0.91$ for $\text{AUC}_{\text{GMRF}} \neq \text{AUC}_{\text{RF}}$, $p = 0.003$ for

503    $\text{RMSE}_{\text{GMRF}} \neq \text{RMSE}_{\text{RF}}$). Compared to GMRFs, RF performance was also more sensitive to

504   withholding data at the edge of the fishery (Fig. 6). The degree to which this was true, however,

505   differed widely between species. For instance, the performance of both models was stable for

506   some species (e.g. darkblotched rockfish in binomial component, blue shark in positive

507   component), indicating that both models captured relevant spatial environment-bycatch

508   relationships in the core area of the fishery and that these relationships remained valid at the edge

509   of the fishery.

510   **Discussion**

511   Our results demonstrate the clear potential of species distribution models to predict fishing

512   activity with higher bycatch rates. While the models' performance varied considerably, even the

513   worst performer (GLM without latitude and longitude) achieved AUC from 0.68-0.86 (Fig. 2)

514   and reduced bycatch-to-target catch ratios on average by 25% for a 5% reduction in fishing effort

515   (Fig. 3). Random forests performed the best, achieving cross-validated AUC above 0.89 for all

516   three West Coast groundfish species (Fig. 2) and reducing bycatch-to-target ratios by 34% for a

517   5% reduction in fishing effort averaged across the six species (Fig. 3). When extrapolating

518   beyond the geographic range of the data, however, RFs were more sensitive to which data were

519   withheld and performed similar to, or worse than, GMRFs (Fig. 6). This is consistent with

520   previous work documenting that more data-driven, complex SDMs can have better interpolation

521   performance but be worse at spatial extrapolation (Heikkinen et al. 2012; Randin et al. 2006).

522   Our results beg the question: if RFs are expected to have higher interpolation accuracy,

523   why ever use GMRFs or GAMs? This is a decidedly relevant concern given the significant

524   research investment in these modeling approaches (Becker et al. 2014; Thorson and Barnett

525   2017). RF will generally have better out-of-sample prediction, largely due to its ability to

526   incorporate nonlinear and interaction effects of covariates inherently (i.e., without user

527   specification). In addition, RF models are much simpler and quicker to both write code for and

528   run. Contrary to references describing RFs as "black boxes" (Prasad et al. 2006; Cutler et al.

529   2007; Elith and Leathwick 2009; Evans and Cushman 2009; Kuhn and Johnson 2013), there are

530   methods for investigating RF model structure, including covariate effects and interactions (Fig.

531   5, Welling et al. 2016). In a similar vein, it is possible to calculate prediction variance and

532   confidence intervals for RF (Fig. 4, Meinshausen 2006; Wager et al. 2014), despite older

533   ecological literature stating otherwise (Cutler et al. 2007; Olden et al. 2008), or using ad-hoc

534   substitutes such as the standard deviation of individual tree predictions (Smoliński and Radtke

535   2017). Additionally, promising theoretical work may soon widen the ability to use RFs for

536   statistical inference by developing asymptotically normal, unbiased point estimates with valid

537   confidence intervals (Mentch and Hooker 2017; Wager and Athey 2017). In many cases, RF's

538   performance advantage is probably sufficient to warrant its use over other semiparametric

539   methods. Yet, there may be several cases where semiparametric methods are preferred.

540   Semiparametric frameworks like GMRFs and GAMs have clear advantages over machine

541   learning algorithms such as RF. First, they can be derived from probability theory and therefore

542   allow for traditional statistical inference on their mean response predictions. GMRF estimates

543   full posterior distributions for the response variable (in this study, probability and/or expected

544   amount of bycatch) everywhere in the spatial domain. This enables us, for instance, to use a

545   GMRF model to identify regions that are above/below a threshold probability (i.e. risk level) of a

546   defined bycatch quantity. By contrast, developing statistical inference based on RF is an active

547   area of research (Biau and Scornet 2016). Thus, GMRF models may be preferred for applications

548   where estimates of model uncertainty are decidedly important, such as using models to produce

549   annual estimates of bycatch (expanded from the observed to unobserved fleet). Second, GMRFs

550   and GAMs explicitly estimate covariate effects with uncertainty intervals, facilitating ecological

551   interpretation of factors significantly affecting the response (Fig. 5). Third, GMRF models are

552   particularly well-suited to incorporate spatial ecological processes, such as movement or spatial

553   variation in mortality, condition, or observation error (Carson and Flemming 2014; Illian et al.

554   2013; Thorson et al. 2017). RFs, on the other hand, are based solely on observations without

555   taking into consideration the data generating process. Fourth, GAMs and GMRFs produce

556   smoothed spatial predictions that are more likely ecologically plausible, whereas RF spatial

557   predictions appear rectangular by default due to splits on the geographical coordinates. This can

558   be mediated, however, by including transformations of the coordinates, fitting a linear model

559   within each node instead of taking the mean (Quinlan 1992, 1993), or including buffer distances

560   as covariates (Hengl et al. 2018). Fifth, RFs can simultaneously have high interpolation accuracy

561   and lower extrapolation accuracy (Fig. 6b). This matters if SDM predictions of bycatch risk are

562   to be used as a spatial management tool, where predictions in areas with sparse or no sampling

563   coverage may be most important. Finally, GMRFs and GAMs can be easily specified to produce

564   independent or exchangeable estimates of a given quantity (e.g., total predicted bycatch in

565   different years), while it is unclear how to assign a specific dependence-structure on factors in a

566   RF model. This is important when using estimates from a spatial model as input in a secondary

567   model, for example, where the secondary model assumes that estimates are independent among

568   years. In situations where these concerns are inconsequential, however, RF are likely the better

569   method for spatial bycatch prediction—they are faster and have better predictive performance

570   than the alternatives. Even when probabilistic conclusions are required, RF may be useful in an

571   exploratory manner given their ease of use, speed, and ability to identify nonlinear and

572   interaction covariate effects for later inclusion in semiparametric models.

573   On a more detailed level, the six species differed widely in their predictability. The West

574   Coast groundfish species were more predictable than the Hawaii longline species, likely because

575   we included several more relevant environmental covariates (WCGOP: sea surface temperature,

576   depth, rocky habitat, in/near RCA, and survey-predicted occurrence; HILL: sea surface

577   temperature) and the WCGOP dataset had 2.5 times the number of observations (WCGOP: 42

578   786; HILL: 16 714). The HILL models' performance could presumably be improved by

579   incorporating more satellite-based environmental covariates capable of explaining the species'

580   distributions, such as chlorophyll and SST-derived frontal indices (Nieto et al. 2017). Among the

581   West Coast species, Pacific halibut was more difficult to predict than the two rockfish species,

582   despite having more positive bycatch occurrences for the models to fit. One possible explanation

583   is that adult halibut move more and have less strict habitat associations than rockfish, decreasing

584   their predictability (Skud 1977; Gunderson 1997). Among the Hawaii longline species,

585   loggerhead turtles were much easier to predict than leatherback turtles and blue sharks, perhaps

586   because they have stronger SST-based habitat preferences where the fishing occurs (Howell et al.

587   2015). Indeed, the loggerhead-temperature association is the basis for TurtleWatch, a decade-

588   long effort to reduce turtle bycatch by providing fishermen with dynamic recommendations of

589   high bycatch risk areas to avoid in the Central Pacific (Howell et al. 2008). Despite plausible life

590   history explanations for some among-species differences in predictability, however, the

591   advantage of fitting time-varying vs. time-constant models appeared to be primarily driven by

592   the species' bycatch rates. In other words, there was little difference in model performance for

593   rare bycatch species (yelloweye rockfish, loggerhead turtle, and leatherback turtle) and greater

594   differences between models for common bycatch species (darkblotched rockfish, Pacific halibut,

595   and blue shark, Figs. 2a and A3).

596   Given their good predictive performance overall, SDMs could be used to support spatial

597   bycatch management, whether static (e.g. design habitat closures to be semi-permanent, such as

598   the Pacific Leatherback Conservation Area along the U.S. West Coast; 50 CFR Part 660) or

599   dynamic in time (e.g. closures change every year, month, week, etc., such as the Loggerhead

600   Turtle Conservation Area along the Southern California Bight, 72FR 31756; Fig. 7, as in Dunn et

601   al. 2016). If so, they should be compared to and integrated with existing tools that aim to reduce

602   bycatch by producing risk maps; examples include Eguchi et al. (2017), TurtleWatch (Howell et

603   al. 2008, 2015), and WhaleWatch (Hazen et al. 2016; Breivik et al. 2016). Both TurtleWatch and

604   WhaleWatch are based on satellite telemetry observations and known habitat preferences of sea

605   turtles and blue whales, in contrast to the models developed here that rely exclusively on

606   fisheries observer data. Fisheries observer datasets cover many more bycatch species than those

607   with satellite tagging programs, which means that SDMs based on fisheries observer data may be

608   more widely applicable than those based solely on satellite telemetry, especially for species with

609   moderate to high bycatch rates. However, we found that SDMs were less effective at predicting

610   rare bycatch events (e.g. yelloweye rockfish, loggerhead turtle, and leatherback turtle in Fig. 2a).

611   Bycatch occurs when non-target species and fishing gear co-occur, both of which are affected by

612   various factors, such as environmental conditions, economics, and behavior (Soykan et al. 2014).

613   Consequently, using fishery observer data combined with animal movement data would provide

614   a comprehensive dataset to develop predictive models of bycatch, and this may be a prudent

615   approach for species with low bycatch rates.

616   Future efforts to use spatial models to predict fisheries bycatch risk should carefully

617   consider the hierarchical structure common to observer datasets with less than 100% coverage. In

618   the typical case where observers are placed on vessels on a trip-by-trip basis and then observe all

619   sets within a trip, the sets are likely not independent (e.g. sets within the same trip, and trips on

620   the same vessel may be correlated). As done in this study, including spatiotemporal correlation

621   structure will account for some of the correlation between sets within a trip because they are

622   presumably closer together in time and space. One approach is to include the nested data

623   structure in the model as random effects (Candy 2004; Thorson et al. 2015*b*), although this can

624   be more complicated than it first appears (e.g. captains and crew transfer between vessels in the

625   Hawaii longline fishery, making it unclear whether including a vessel effect is appropriate).

626   Roberts et al. (2017) make an excellent case for an alternative approach, using block cross-

627   validation to account for spatial, temporal, and group dependence structure when validating and

628   selecting models. A related issue with the delta-models used here is that they assume two

629   independent processes determine the probability and amount of bycatch (i.e., the binomial and

630   positive components of the delta-model). Thorson (2018) and Cantoni et al. (2017) both recently

631   demonstrated that this is unlikely to be true, and proposed solutions that include parameters

632   allowing for dependence between the binomial and positive components. Another relatively

633   simple approach is to use the compound Poisson-gamma, or Tweedie distribution, which Stock et

634   al. (2019) used in a spatial bycatch analysis and outperformed the equivalent delta-GAM. All of

635   these solutions should improve spatiotemporal bycatch predictions in the future.

636   Finally, just as single-species fisheries management paints a rosier picture than can truly

637   be implemented, the results presented here are unrealistic in their treatment of multispecies

638   fisheries because bycatch prediction cannot be optimized for each individual species

639   simultaneously. These results are still useful if fisheries managers are particularly concerned

640   about a single species, but less so if reducing bycatch of multiple species is the objective. Both

641   RF and GMRF models have multivariate extensions that could fruitfully be applied to

642   multispecies spatial bycatch prediction, and future work should investigate this possibility

643   (Thorson et al. 2015a; Ishwaran and Kogalur 2018; Thorson and Barnett 2017).

644

653

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. 233 p.

Aguirre-Gutiérrez, J., Carvalheiro, L. G., Polce, C., van Loon, E. E., Raes, N., Reemer, M., and Biesmeijer, J. C. 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria - Dutch hoverflies as a case study. PLoS ONE 8:e63708.

Araújo, M.B. and Rahbek, C. 2006. How does climate change affect biodiversity? Science 313:1396-1397.

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3:327–358.

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. Endangered Species Research 23:1–22.

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.

Benson, S. R., Eguchi, T., Foley, D. G., Forney, K. A., Bailey, H., Hitipeuw, C., Samber, B. P., Tapilatu, R. F., Rei, V., Ramohia, P., and Pita, J. 2011. Large-scale movements and high-use areas of western Pacific leatherback turtles, *Dermochelys coriacea*. Ecosphere 2(7):1-27.

Biau, G., and Scornet, E. 2016. A random forest guided tour. TEST 25:197–227.

Bjørnstad, O. N., and Falck, W. 2001. Nonparametric spatial covariance functions: estimation and testing. Environmental and Ecological Statistics 8:53-70.

Boyce, J. R. 1996. An economic analysis of the fisheries bycatch problem. Journal of Environmental Economics and Management 31(3):314-336.

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. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-114, 323p.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

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

Breivik, O. N., Storvik, G., and Nedreaas, K. 2016. Latent Gaussian models to decide on spatial

687   closures for bycatch management in the Barents Sea shrimp fishery. Canadian Journal of
688   Fisheries and Aquatic Science 73:1271–1280.

689 Breivik, O. N., G. Storvik, and K. Nedreaas. 2017. Latent Gaussian models to predict historical
690   bycatch in commercial fishery. Fisheries Research 185:62-72.

691 Candy, S. G. 2004. Modelling catch and effort data using generalised linear models, the Tweedie
692   distribution, random vessel effects and random stratum-by-year effects. CCAMLR Science
693   11:59-80.

694 Cantoni, E., J. M. Flemming, and A. H. Welsh. 2017. A random-effects hurdle model for
695   predicting bycatch of endangered marine species. The Annals of Applied Statistics
696   11(4):2178-2199.

697 Carretta, J. V., J. E. Moore, and K. A. Forney. 2017. Regression tree and ratio estimates of
698   marine mammal, sea turtle, and seabird bycatch in the California drift gillnet fishery: 1990-
699   2015. NOAA Tech. Memo., NOAA-TM-NMFS-SWFSC-568. 83 p.

700 Carson, S., and J. M. Flemming. 2014. Seal encounters at sea: A contemporary spatial approach
701   using R-INLA. Ecological Modelling 291:175–181.

702 Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: synthetic
703   minority over-sampling technique. Journal of Artificial Intelligence Research 16:321–357.

704 Conn, P.B., D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. 2015.
705   Using spatiaotemporal statistical models to estimate animal abundance and infer ecological
706   dynamids from survey counts. Ecological Monographs 85(2):235–252.

707 Cosandey-Godin, A., E. T. Krainski, B. Worm, and J. M. Flemming. 2015. Applying Bayesian
708   spatiotemporal models to fisheries bycatch in the Canadian Arctic. Canadian Journal of
709   Fisheries and Aquatic Sciences 72:186–197.

710 Cutler, D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler.
711   2007. Random forests for classification in ecology. Ecology 88:2783–2792.

712 Davies, R. W. D., S. J. Cripps, A. Nickson, and G. Porter. 2009. Defining and estimating global
713   marine fisheries bycatch. Marine Policy 33(4):661-672.

714 Diggle, P. J., J. A. Tawn, and R. A. Moyeed. 1998. Model-based geostatistics. Journal of the
715   Royal Statistical Society: Series C (Applied Statistics) 47(3):299-350.

716 Dormann, C. F., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A.
717   Hirzel, W. Jetz, W. D. Kissling, I. Kühn, R. Ohlemuller, P. R. Peres-Neto, B. Reineking, B.
718   Schroder, F. M. Schurr, and R. Wilson 2007. Methods to account for spatial autocorrelation
719   in the analysis of species distributional data: a review. Ecography 30(5):609-628.

720 Drake, J. M., C. Randin, and A. Guisan. 2006. Modelling ecological niches with support vector
721   machines. Journal of Applied Ecology 43:424–432.

722 Dunn, D. C., S. M. Maxwell, A. M. Boustany, and P. N. Halpin. 2016. Dynamic ocean
723   management increases the efficiency and efficacy of fisheries management. Proceedings of
724   the National Academy of Sciences 113:668–673.

725 Eguchi, T., S. R. Benson, D. G. Foley, and K. A. Forney. 2017. Predicting overlap between drift
726   gillnet fishing and leatherback turtle habitat in the California Current Ecosystem. Fisheries
727   Oceanography 26(1):17-33.

728 Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J.,
729   Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A.,
730   Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T.,
731   Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberon, J., Williams,
732   S., Wisz, M. S., and Zimmermann, N. E. 2006. Novel methods improve prediction of
733   species' distributions from occurrence data. Ecography 29:129–151.

734 Elith, J., and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding
735   reasons for differing performances of species distribution models. Ecography 32(1):66-77.

736 Elith, J., and Leathwick, J. R. 2009. Species distribution models: ecological explanation and
737   prediction across space and time. Annual Review of Ecology, Evolution, and Systematics
738   40:677–697.

739 Evans, J. S., and Cushman, S. A. 2009. Gradient modeling of conifer species using random
740   forests. Landscape Ecology 24(5):673-683.

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

743 FAO. 1995. Code of Conduct for Responsible Fisheries. Rome, FAO. 41 p.

744 Gilman, E., Kobayashi, D., Swenarton, T., Brothers, N., Dalzell, P., and Kinan-Kelly, I. 2007.
745   Reducing sea turtle interactions in the Hawaii-based longline swordfish fishery. Biological
746   Conservation 139:19–28.

747 Golding, N., and Purse, B. V. 2016. Fast and flexible Bayesian species distribution modelling
748   using Gaussian processes. Methods in Ecology and Evolution 7(5): 598-608.

749 Guélat, J., and Kéry, M. 2018. Effects of spatial autocorrelation and imperfect detection on
750   species distribution models. Methods in Ecology and Evolution 9(6):1614-1625.

751 Guisan, A., and Thuiller, W. 2005. Predicting species distribution: offering more than simple
752   habitat models. Ecology Letters 8:993–1009.

753 Gunderson, D. R. 1997. Spatial patterns in the dynamics of slope rockfish stocks and their
754   implications for management. Fishery Bulletin 95:219–230.

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

759 Hand, D.J. 2009. Measuring classifier performance: a coherent alternative to the area under the
760   ROC curve. Machine Learning 77(1):103-123.

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

763 Hazen, E. L., Palacios, D. M., Forney, K. A., Howell, E. A., Becker, E., Hoover, A. L., Irvine, L.,
764   Deangelis, M., Bograd, S. J., Mate, B. R., and Bailey, H. 2016. WhaleWatch: a dynamic
765   management tool for predicting blue whale density in the California Current. Journal of
766   Applied Ecology 54(5):1415-1428.

767 Heikkinen, R. K., Marmion, M., and Luoto, M. 2012. Does the interpolation accuracy of species
768   distribution models come at the expense of transferability? Ecography 35(3):276-288.

769 Hengl, T., Nussbaum, M., Wright, M. N., and G. B. Heuvelink. 2018. Random Forest as a
770   generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ
771   Preprints e26693v1.

772 Hooten, M. B., and Hobbs, N. T. 2015. A guide to Bayesian model selection for ecologists.
773   Ecology 85:3–28.

774 Howell, E. A., Hoover, A., Benson, S. R., Bailey, H., Polovina, J. J., Seminoff, J. A., and Dutton,
775   P. H. 2015. Enhancing the TurtleWatch product for leatherback sea turtles, a dynamic
776   habitat model for ecosystem-based management. Fisheries Oceanography 24:57–68.

777 Howell, E. A., Kobayashi, D. R., Parker, D. M., Balazs, G. H., and Polovina, J. J. 2008.
778   TurtleWatch: a tool to aid in the bycatch reduction of loggerhead turtles *Caretta caretta* in
779   the Hawaii-based pelagic longline fishery. Endangered Species Research 5:267–278.

780 Ianelli, J. N., and Stram, D. L. 2015. Estimating impacts of the pollock fishery bycatch on
781   western Alaska Chinook salmon. ICES Journal of Marine Science 72:1159–1172.

782 Illian, J. B., Martino, S., Sørbye, S. H., Gallego-Fernández, J. B., Zunzunegui, M., Esquivias, M.
783   P., and Travis, J. M. J. 2013. Fitting complex ecological point process models with
784   integrated nested Laplace approximation. Methods in Ecology and Evolution 4:305–315.

785 Ishwaran, H., and Kogalur, U. B. 2018. Random forests for survival, regression and classification
786   (RF-SRC). R package version 2.6.1.

787 Kammann, E. E., and Wand, M. P. 2003. Geoadditive models. Journal of the Royal Statistical
788   Society: Series C (Applied Statistics) 52(1):1-18.

789 Kelleher, K. 2005. Discards in the world's marine fisheries: An update. FAO Fisheries Technical
790   Paper. No. 470. Rome, FAO. 131p.

791 Kneib, T., Müller, J., and Hothorn, T. 2008. Spatial smoothing techniques for the assessment of
792   habitat suitability. Environmental and Ecological Statistics 15(3):343-364.

793 Kobayashi, D. R., Polovina, J. J., Parker, D. M., Kamezaki, N., Cheng, I. J., Uchida, I., Dutton, P.
794   H., and Balazs, G. H. 2008. Pelagic habitat characterization of loggerhead sea turtles,
795   *Caretta caretta*, in the North Pacific Ocean (1997–2006): insights from satellite tag tracking
796   and remotely sensed data. Journal of Experimental Marine Biology and Ecology 356(1):96-
797   114.

798 Kuhn, M., and Johnson, K. 2013. Applied predictive modeling. Springer Science & Business
799   Media, New York, NY.

800 Leathwick, J. R., Elith, J., and Hastie, T. 2006. Comparative performance of generalized additive
801   models and multivariate adaptive regression splines for statistical modelling of species
802   distributions. Ecological Modelling 199:188–196.

803 Lecomte, J. B., Benoît, H. P., Ancelet, S., Etienne, M. P., Bel, L., and Parent, E. 2013. Compound
804   Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable
805   sampling volume. Methods in Ecology and Evolution 4(12):1159-1166.

806 Lewison, R., Hobday, A. J., Maxwell, S., Hazen, E., Hartog, J. R., Dunn, D. C., Briscoe, D.,
807   Fossette, S., O'Keefe, C. E., Barnes, M., Abecassis, M., Bograd, S., Bethoney, N. D.,
808   Bailey, H., Wiley, D., Andrews, S., Hazen, L., and Crowder, L. B. 2015. Dynamic ocean
809   management: Identifying the critical ingredients of dynamic approaches to ocean resource
810   management. BioScience 65:486–498.

811 Li, S., and Pan, M. 2011. Fishing opportunities under the sea turtle interaction caps—a spatial
812   bio-economic model for Hawaii-based longline swordfish fishery. SOEST 11-02. Joint
813   Institute for Marine and Atmospheric Research (JIMAR) Contribution 11-378.

814 Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. R News 2:18–22.

815 Liggins, G. W., Bradley, M. J., and Kennelly, S. J. 1997. Detection of bias in observer-based
816   estimates of retained and discarded catches from a multi species trawl fishery. Fisheries
817   Research 32(2):133-147.

818 Lindgren, F., and Rue, H. 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical
819   Software 63:1–25.

820 Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and

Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society. Series B: Statistical Methodology 73:423–498.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., and Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions 15(1):59-69.

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. Ecological Applications 25(2):416-429.

Maunder, M. N., and Punt, A. E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries Research 70(2):141-159.

McCracken, M. L. 2004. Modeling a very rare event to estimate sea turtle bycatch: Lessons learned. U.S. Dep. Commer., NOAA Tech. Memo., NOAA-TM-NMFS-PIFSC-3. 30p.

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. Journal of Computational and Graphical Statistics 26(3):589-597.

Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E., and Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography 37(12):1267-1281.

(NMFS) National Marine Fisheries Service. 2013. Groundfish essential fish habitat synthesis: a report to the Pacific Fishery Management Council. Seattle, WA.

(NMFS) National Marine Fisheries Service. 2016. U.S. National Bycatch Report First Edition Update 2 [Benaka, L. R., Bullock, D., Davis, J., Seney, E. E., and Winarsoo, H., Editors]. U.S. Dep. Commer. 90 p.

(NWFSC) Northwest Fisheries Science Center. 2006. Observer coverage plan: Sampling plan and logistics for the West Coast Groundfish Observer Program. NOAA, West Coast Groundfish Observer Program, 2725 Montlake Blvd E, Seattle, WA. 10p. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.231.1675.

(NWFSC) Northwest Fisheries Science Center. 2016. West Coast Groundfish Observer Program 2016 catch share training manual. Seattle, WA.

Nichols, W. J., Resendiz, A., Seminoff, J. A., and Resendiz, B. 2000. Transpacific migration of a loggerhead turtle monitored by satellite telemetry. Bulletin of Marine Science 67(3):937-947.

854　Nieto, K., Xu, Y., Teo, S. L. H., McClatchie, S., and Holmes, J. 2017. How important are coastal
855　　　fronts to albacore tuna (Thunnus alalunga) habitat in the Northeast Pacific Ocean? Progress
856　　　in Oceanography 150:62–71.

857　NOAA Fisheries West Coast Region. 2015. Rockfish conservation areas.
858　　　http://www.westcoast.fisheries.noaa.gov/fisheries/management/groundfish_closures/rockfis
859　　　h_areas.html.

860　NOAA National Centers for Environmental Information. 2015. U.S. Coastal Relief Model.
861　　　https://www.ngdc.noaa.gov/mgg/coastal/crm.html.

862　Olden, J. D., Lawler, J. J., and Poff, N. L. 2008. Machine learning methods without tears: a
863　　　primer for ecologists. The Quarterly Review of Biology 83:171–193.

864　Parmesan, C., and Yohe, G. 2003. A globally coherent fingerprint of climate change impacts
865　　　across natural systems. Nature 421:37–42.

866　Pearce, J., and Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species
867　　　distribution models using logistic regression. Ecological Modelling 128:127–147.

868　Pennington, M. 1983. Efficient estimators of abundance, for fish and plankton surveys.
869　　　Biometrics 39:281–286.

870　Péron, G., Ferrand, Y., Gossmann, F., Bastat, C., Guenezan, M., and Gimenez, O. 2011.
871　　　Nonparametric spatial regression of survival probability: visualization of population sinks in
872　　　Eurasian Woodcock. Ecology 92(8):1672-1679.

873　Phillips, S. J., Anderson, R. P., and Schapire, R. E. 2006. Maximum entropy modeling of species
874　　　geographic distributions. Ecological Modelling 190:231–259.

875　Phillips, S. J., and Dudík, M. 2008. Modeling of species distribution with Maxent: new
876　　　extensions and a comprehensive evalutation. Ecography 31:161–175.

877　(PIROP) Pacific Islands Regional Observer Program. 2014. Hawaii longline observer program
878　　　observer field manual. Version LM.14.04. NOAA/IRC 1, 1845 Wasp Blvd. Bldg. 176,
879　　　Honolulu, Hawaii 96818.

880　Pons, M., Marroni, S., Machado, I., Ghattas, B., and Domingo, A. 2009. Machine learning
881　　　procedures: an application to by-catch data of the marine turtles *Caretta Caretta* in the
882　　　Southwestern Atlantic Ocean. Collect. Vol. Sci. Pap. ICCAT. 64(7):2443-2454.

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

885　Prasad, A. M., Iverson, L. R., and Liaw, A. 2006. Newer classification and regression tree
886　　　techniques: bagging and random forests for ecological prediction. Ecosystems 9(2):181-

887 199.

888 Quinlan, J. R. 1992. Learning with continuous classes. *In* 5th Australian Joint Conference on
889 Artificial Intelligence, pp. 343–348.

890 Quinlan, J. R. 1993. Combining instance-based and model-based learning. *In* Proceedings of the
891 Tenth International Conference on Machine Learning, pp. 236–243.

892 R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for
893 Statistical Computing, Vienna, Austria.

894 Randin, C. F., Dirnböck, T., Dullinger, S., Zimmermann, N. E., Zappa, M., and Guisan, A. 2006.
895 Are niche-based species distribution models transferable in space? Journal of Biogeography
896 33(10):1689-1703.

897 Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. 2007.
898 Daily high-resolution-blended analyses for sea surface temperature. Journal of Climate
899 20:5473–5496.

900 Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S.,
901 Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and
902 Dormann, C. F. 2017. Cross-validation strategies for data with temporal, spatial,
903 hierarchical, or phylogenetic structure. Ecography 40(8):913-929.

904 Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian
905 models by using integrated nested Laplace approximations. Journal of the Royal Statistical
906 Society. Series B: Statistical Methodology 71:319–392.

907 Scales, K. L., Miller, P. I., Ingram, S. N., Hazen, E. L., Bograd, S. J., and Phillips, R. A. 2016.
908 Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble
909 ecological niche models. Diversity and Distributions 22:212-224.

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

913 Shmueli, G. 2010. To explain or to predict? Statistical Science 25:289–310.

914 Skud, B. E. 1977. Drift, migration, and intermingling of Pacific halibut stocks. Scientific Report
915 No. 63. International Pacific Halibut Commission, Seattle, WA.

916 Smoliński, S., and Radtke, K. 2017. Spatial prediction of demersal fish diversity in the Baltic
917 Sea: comparison of machine learning and regression-based techniques. ICES Journal of
918 Marine Science 74:102–111.

919 Soykan, C. U., Eguchi, T., Kohin, S., and Dewar, H. 2014. Prediction of fishing effort

920   distributions using boosted regression trees. Ecological Applications 24(1):71-83.

921   Stefánsson, G. 1996. Analysis of groundfish survey abundance data: combining the GLM and
922   delta approaches. ICES Journal of Marine Science 53(3):577–588.

923   Stock, B. C., Ward, E. J., Thorson, J. T., Jannot, J. E., and Semmens, B. X. 2019. The utility of
924   spatial model-based estimators of unobserved bycatch. ICES Journal of Marine Science
925   76:255–267.

926   Sumaila, U. R., Cheung, W. W. L., Lam, V. W. Y., Pauly, D., and Herrick, S. 2011. Climate
927   change impacts on the biophysics and economics of world fisheries. Nature Climate Change
928   1:449–456.

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

932   Thorson, J. T., Fonner, R., Haltuch, M. A., Ono, K., and Winker, H. 2017. Accounting for
933   spatiotemporal variation and fisher targeting when estimating abundance from multispecies
934   fishery data. Canadian Journal of Fisheries and Aquatic Sciences 74:1794–1807.

935   Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., and Kristensen, K.
936   2015*a*. Spatial factor analysis: a new tool for estimating joint species distributions and
937   correlations in species range. Methods in Ecology and Evolution 6:627–637.

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

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

944   Torgo, L. 2010. Data mining with R: Learning with case studies. Chapman and Hall/CRC, Boca
945   Raton, FL.

946   Venables, W. N., and Dichmont, C. M. 2004. GLMs, GAMs and GLMMs: an overview of theory
947   for applications in fisheries research. Fisheries Research 70(2):319-337.

948   Wager, S., and Athey, S. 2017. Estimation and inference of heterogeneous treatment effects using
949   random forests. Journal of the American Statistical Association.

950   Wager, S., Hastie, T., and Efron, B. 2014. Confidence intervals for random forests: the jackknife
951   and the infinitesimal jackknife. Journal of Machine Learning Research 15:1625–1651.

952   Ward, E. J., Holmes, E. E., Thorson, J. T., and Collen, B. 2014. Complexity is costly: a meta-

953   analysis of parametric and non-parametric methods for short-term population forecasting.
954   Oikos 123:652–661.

955 Ward, E. J., Jannot, J. E., Lee, Y. W., Ono, K., Shelton, A. O., and Thorson, J. T. 2015. Using
956   spatiotemporal species distribution models to identify temporally evolving hotspots of
957   species co-occurrence. Ecological Applications 25:2198–2209.

958 Watson, J. T., Essington, T. E., Lennert-Cody, C. E., and Hall, M. A. 2009. Trade-offs in the
959   design of fishery closures: Management of silky shark bycatch in the Eastern Pacific Ocean
960   Tuna fishery. Conservation Biology 23:626–635.

961 Welling, S. H., Refsgaard, H. H. F., Brockhoff, P. B., and Clemmensen, L. H. 2016. Forest floor
962   visualizations of random forests. ArXiv e-prints. http://arxiv.org/abs/1605.09196.

963 Wood, S. N. 2017. Generalized additive models: An introduction with R (2nd ed.). Chapman &
964   Hall/CRC, Boca Raton, FL.

965 Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation
966   of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series
967   B (Statistical Methodology) 73(1):3-36.

968

969 **Data Accessibility**

970 Unsummarized fisheries observer data are deemed confidential. Code to download and process

971 publicly available fisheries survey data, run each of the models, and replicate figures is provided

972 at https://github.com/brianstock/spatial-bycatch. Summarized reports of the fisheries observer

973 program data are available at:

974 https://www.nwfsc.noaa.gov/research/divisions/fram/observation/data_products/data_library.cfm

975 (U.S. West Coast groundfish) and http://www.fpir.noaa.gov/OBS/obs_hi_ll_ss_rprts.html

976 (Hawaii longline).

977

978 **Author contributions**

979 BCS, EW, TE, and BXS conceived of and designed the study with input from JJ, JT, and BF;

980 BCS, JJ, and BF curated the data; BCS performed the analysis with input from EW and BXS;

981 BCS, EW, and JT drafted the manuscript; all authors provided critical reviews.

982

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 forests (RF)	Nonparametric			
RF-BASE		Low	randomForest	+ Lat + Lon
RF-DOWN		Low	randomForest	+ Lat + Lon
RF-SMOTE		Low	caret	+ Lat + Lon

983

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

985 fishing locations. The GLM model serves as the baseline model—no spatial data included. GAM

986 models fit 2-d splines on geographical coordinates (i.e. latitude and longitude), either constant

987 across years (GAM-CONSTANT) or estimating a different spline for each year (GAM-YEAR).

988 Gaussian Markov random field (GMRF) models incorporate spatial locations by estimating the

989 covariance between locations as a random field (with stationary Matern covariance function). As

990 for GAMs, we fit GMRFs that estimate one random field kept constant across years (GMRF-

991 CONSTANT) or estimate a random field for each year (GMRF-YEAR). RF is nonparametric and

992 thus only incorporates spatial locations by including covariates of latitude and longitude. All

993 models for a given species were fit using the same non-spatial covariates (habitat, depth, SST,

994 etc.). We used the R packages 'mgcv' (GLM and GAM), 'INLA' (GMRF), 'randomForest' (RF),

995 and 'caret' (RF-SMOTE).

996   Figure 1. Spatial extent of the two fisheries observer datasets. a) Fishing effort in the West Coast

997   groundfish trawl fishery from 2003 to 2012 (42,786 haul locations). b) Fishing effort in the

998   shallow-set Hawaii longline swordfish fishery from 1994 to 2014 (16,714 set locations).

999   Bivariate kernel density estimates of fishing effort were used to smooth the data ('bkde2D'

1000   function in R package 'KernSmooth').

1001

1002   Figure 2. Predictive performance boxplots of the a) binomial and b) positive components of the

1003   delta-model on test data from 5-fold cross-validation repeated 10 times: a) AUC for the binomial

1004   component, and b) normalized RMSE for the positive component. Across species, random

1005   forests (RFs) outperformed GAMs and GMRFs (highest AUC, lowest RMSE). Significant ($p <$

1006 0.05, Tukey's HSD) within-species performance differences from RF and GMRF are denoted

1007   with black and blue asterisks, respectively. Only the best submodel, e.g. CONSTANT or YEAR,

1008   within each model class for each species is shown here (see Supplement). Species abbreviations:

1009   DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye rockfish, LOGG =

1010   loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark.

1011

1012   Figure 3. Bycatch-to-target species catch ratio achieved by using the binomial component of the

1013   delta-model to predict and remove fishing sets in the test data, relative to the bycatch-to-target

1014   ratio with no fishing sets removed. Lines show median of 50 cross-validation runs for each

1015   model class (5-fold CV repeated 10 times), averaged across the six species. Shaded areas are

1016   bootstrapped 95% confidence intervals for the median. Random forest (RF) performed the best,

1017   reducing the bycatch-to-target ratio by 34% when removing 5% of fishing, and by 50% when

1018   removing 10% of fishing. As in Figure 2, only the best submodel within each model class (e.g.

1019 CONSTANT or YEAR) for each species is shown here.

1020

1021 Figure 4. Maps of predicted blue shark bycatch density with uncertainty for the Hawaii longline

1022 swordfish fishery in 2014. Left panels show the mean bycatch density, log(number per set),

1023 estimated using the four model frameworks: a) GLM, b) GAM, c) GMRF, and d) RF. Right

1024 panels show the log-variance of bycatch density: e) GLM, f) GAM, g) GMRF, and h) RF. All

1025 models predict higher, and more variable, bycatch density in the northwest area of the fishery.

1026 Maps created by GMRF and RF show artifacts of their construction: the mesh triangulation is

1027 evident in the GMRF variance map (g), and the sharp gradients in the RF mean (d) and variance

1028 (h) maps are a consequence of RF trees splitting on latitude and longitude.

1029

1030 Figure 5. Covariate effects on the probability of darkblotched rockfish bycatch (binomial

1031 component of the delta-model). All models estimate a positive effect of PredOcc (predicted

1032 occurrence from survey data, left column), quadratic effect of Depth (center), and positive effect

1033 of In/near RCA (haul location inside or near rockfish conservation area boundary, right). GLM,

1034 GAM, and GMRF covariate effects are marginal posterior distributions ('mgcv' and 'INLA'

1035 packages in R), and RF covariate effects are feature contributions ('forestFloor' package in R).

1036

1037 Figure 6. Predictive performance of the a) binomial and b) positive components of the delta-

1038 models at test locations beyond the geographic data range (i.e. spatial extrapolation). We fit a 2d

1039 kernel density estimate at each observed fishing location ('bkde2D' function in 'KernSmooth' R

1040 package), then sequentially used the lowest 0.5%, 1%, 2%, 5%, 10%, and 20% density locations

1041 as test datasets. Triangles show median model performance from 5-fold cross-validation runs

1042   with random test/train splits (Fig. 2). When extrapolating spatially, all models performed equal to

1043   or worse than when interpolating (i.e. points are lower AUC and higher RMSE than triangles at

1044   20% removed). Compared to GMRF, RF performance was more sensitive to withholding data at

1045   the edge of the fishery (i.e. regression lines have steeper slopes). Missing points and lines

1046   indicate the model failed to converge, as for GMRF with yelloweye rockfish. Species

1047   abbreviations: DBRK = darkblotched rockfish, PHLB = Pacific halibut, YEYE = yelloweye

1048   rockfish, LOGG = loggerhead turtle, LEATH = leatherback turtle, BLUE = blue shark.

1049

1050   Figure 7. GMRF-YEAR random field for bycatch probability of darkblotched rockfish from

1051   2008 to 2012.

1052

1053   Additional supplemental items may be found in the online version of this article:

1054   Supplemental Figures S1-S18. Maps of model residuals for all species for the binomial and

1055   positive components of the delta-model.

1056   Figure A1. Binomial component predictive performance (AUC) for the three random forest

1057   (RF) submodels for all six species.

1058   Figure A2. Binomial component predictive performance (AUC) for the two GMRF models:

1059   CONSTANT (white, one random field constant across years) and YEAR (grey, random

1060   field fit for each year).

1061   Figure A3. GMRF-CONSTANT random field for bycatch probability of the three U.S. West

1062   Coast groundfish species (DBRK = darkblotched rockfish, PHLB = Pacific halibut, and

1063   YEYE = yelloweye rockfish).

1064   Figure A4. Bycatch-to-target species catch ratio achieved for each species by using the

1065   binomial component of each delta-model to predict and remove fishing sets in the test

1066   data, relative to the bycatch-to-target ratio with no fishing sets removed.

1067   Figure A5. Distribution of blue shark bycatch by latitude.

1068   Figure A6. Spatial spline correlograms of residuals from the A) binomial and B) positive

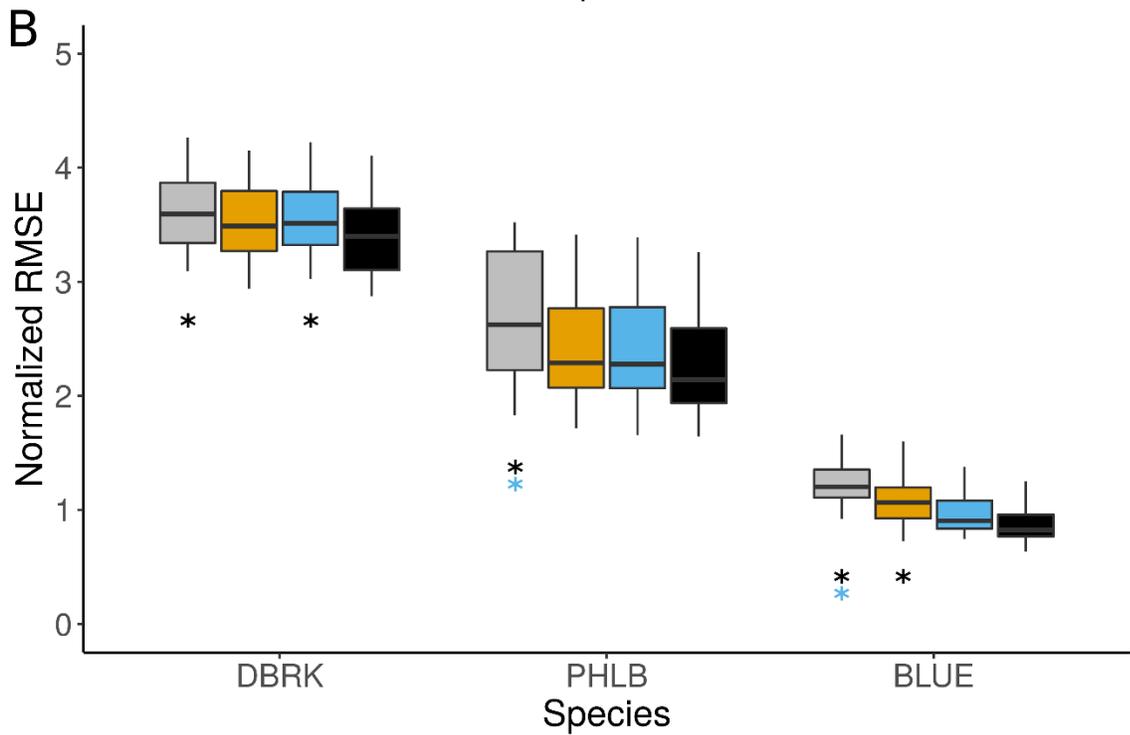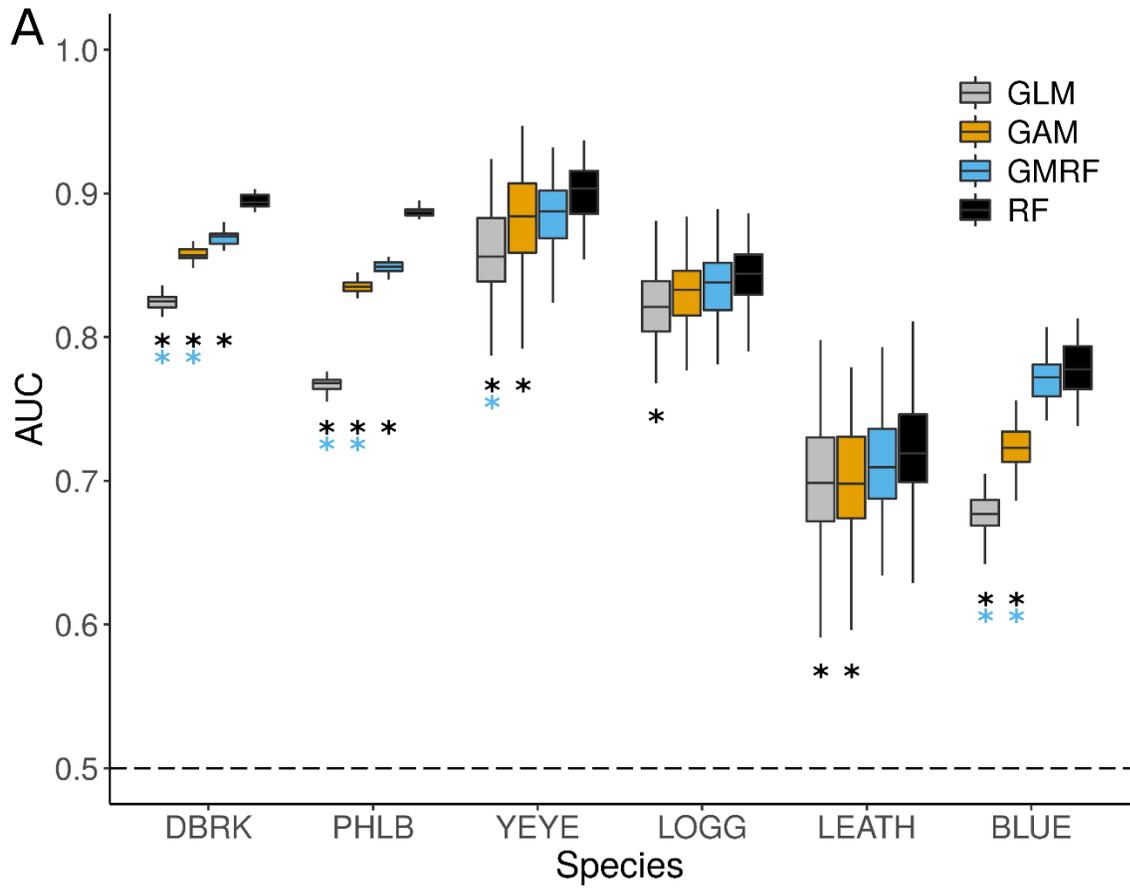1069   components of the delta-models.
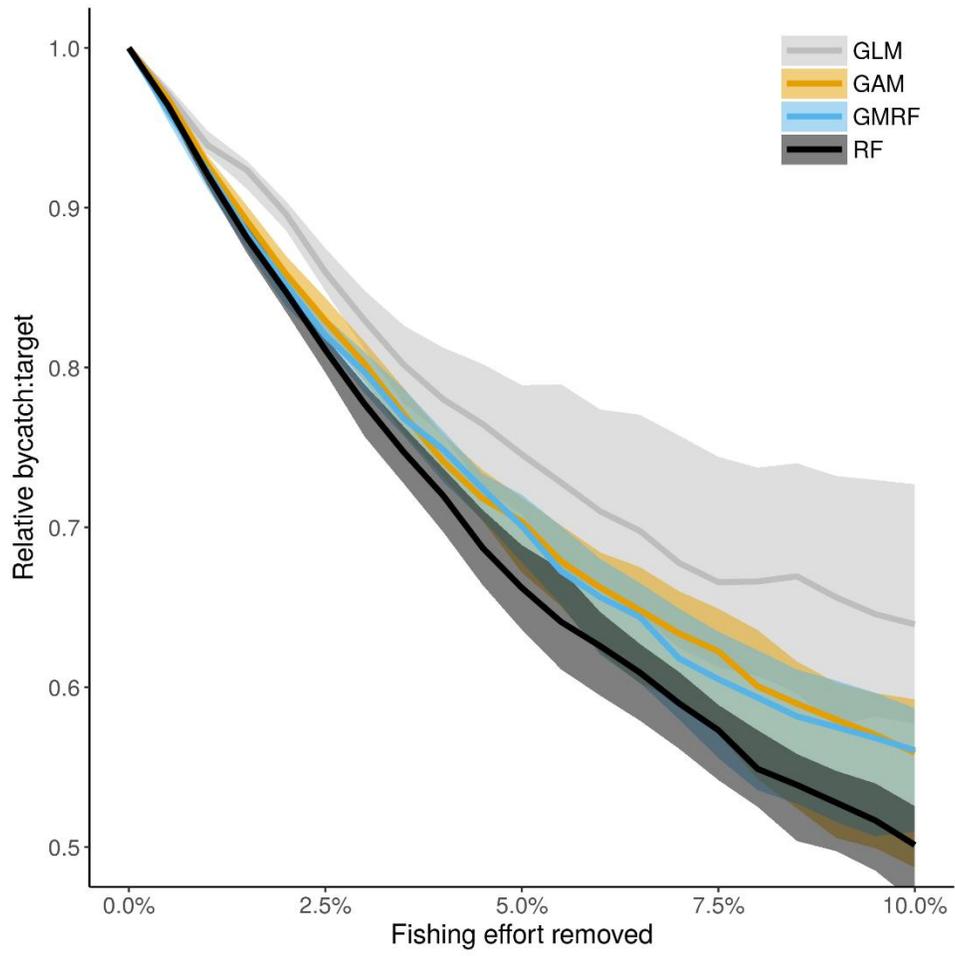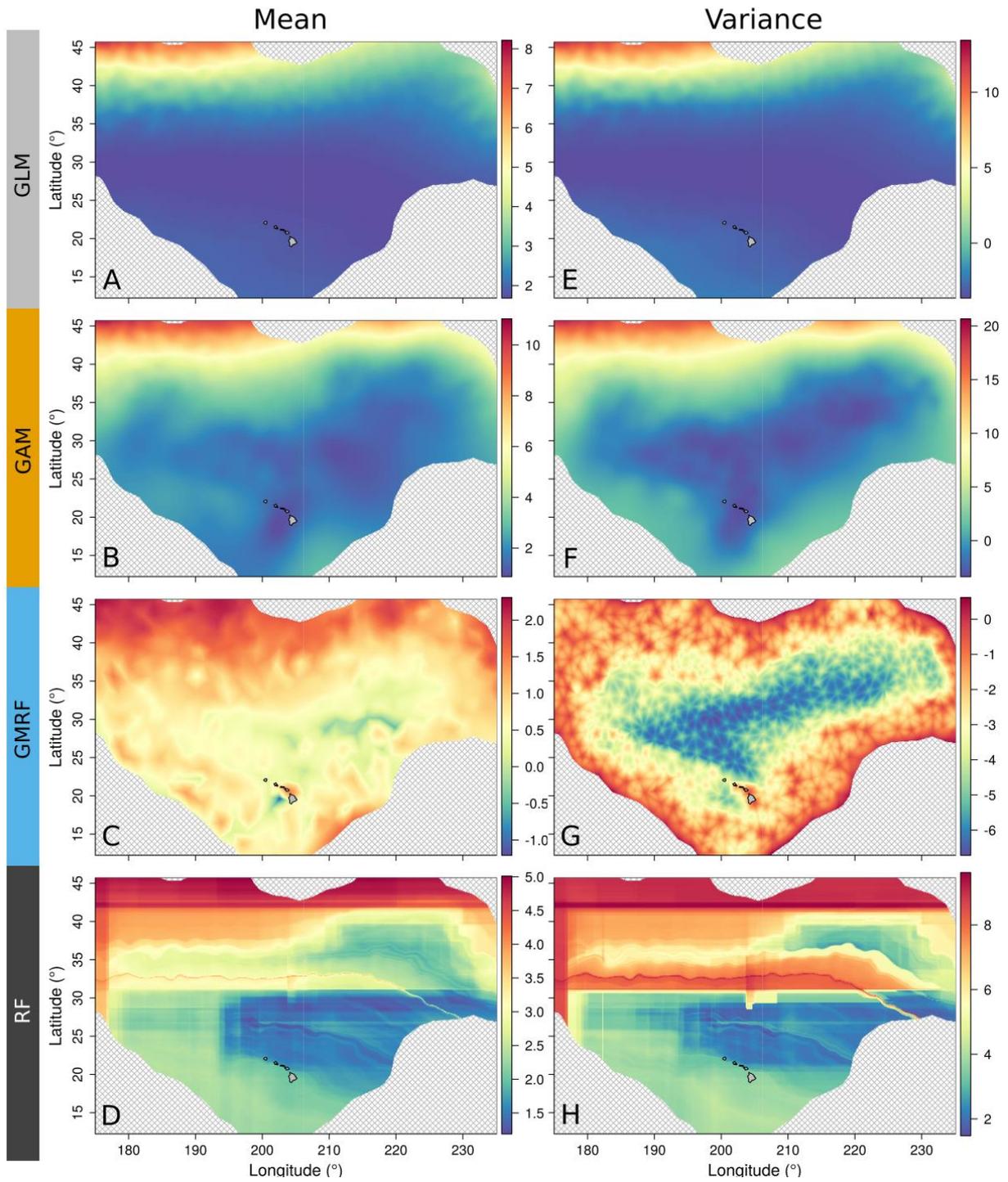
1070

1071   Figure 1.
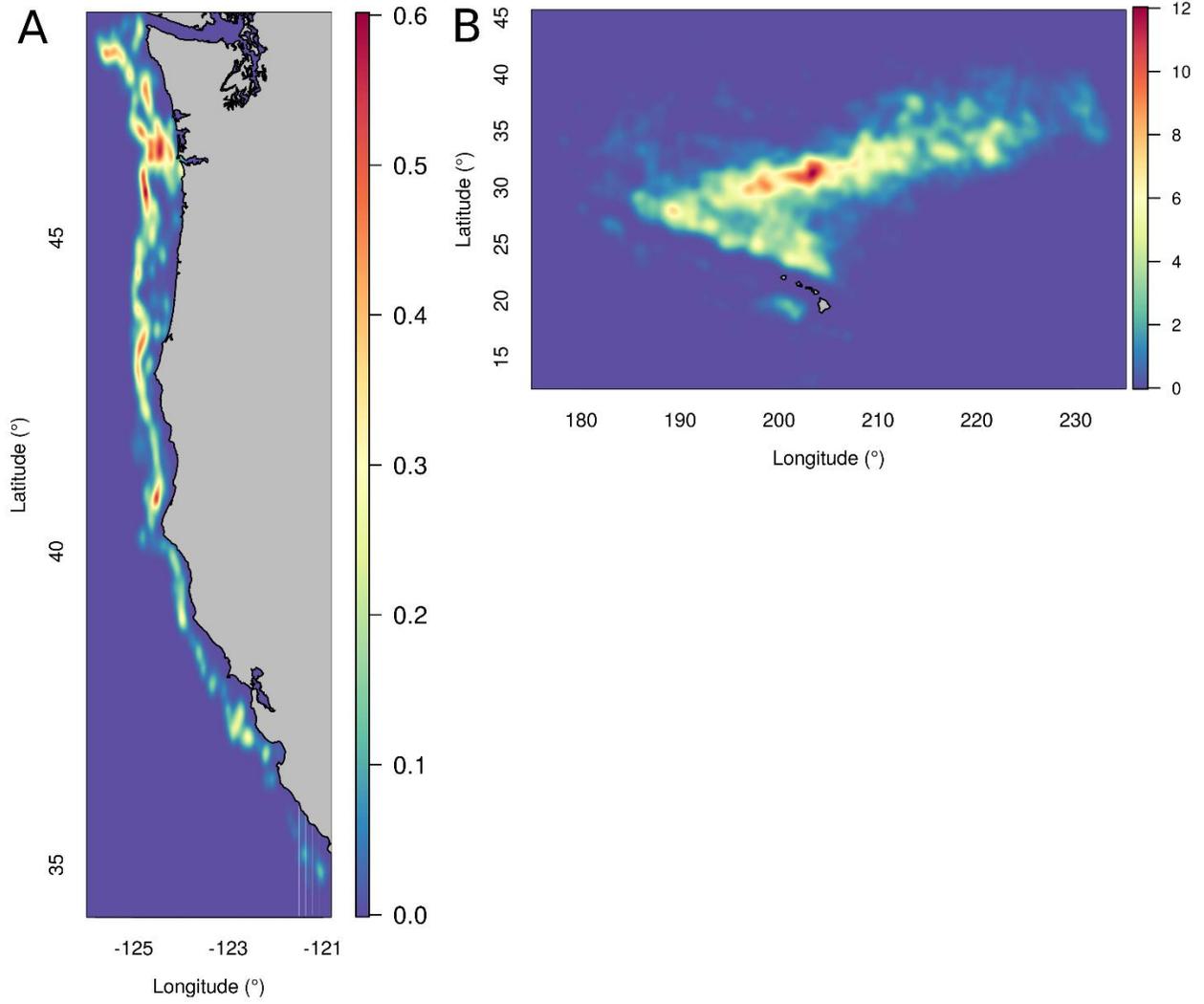
1072

1073

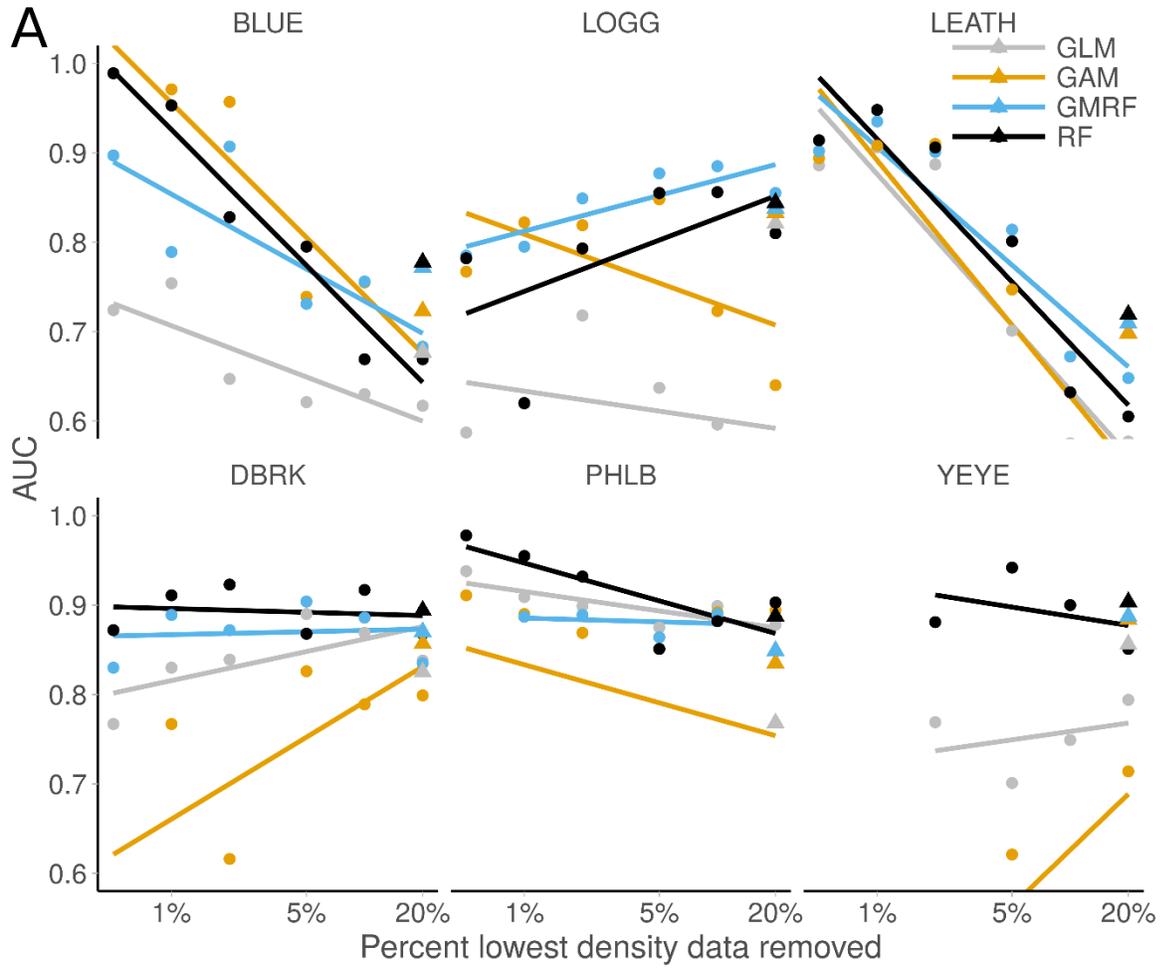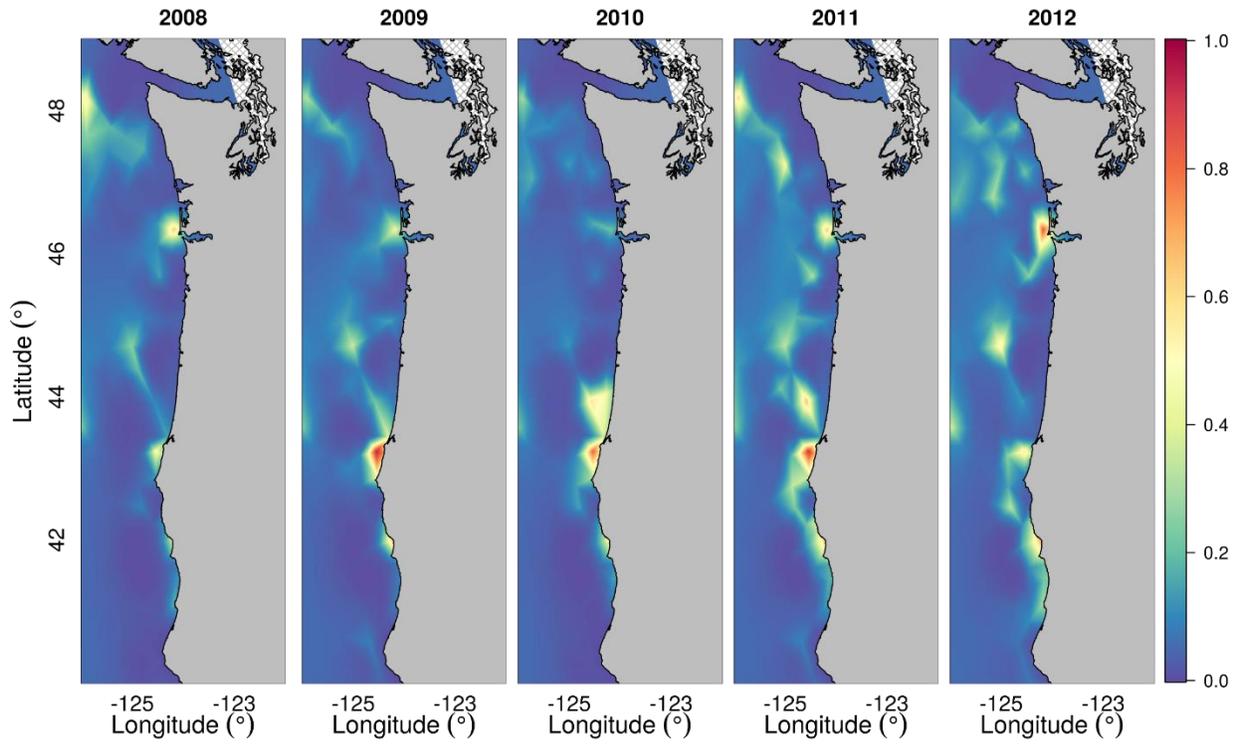Figure 2.

Figure 3.

1081   Figure 4.

1082

1083

1084   Figure 5.
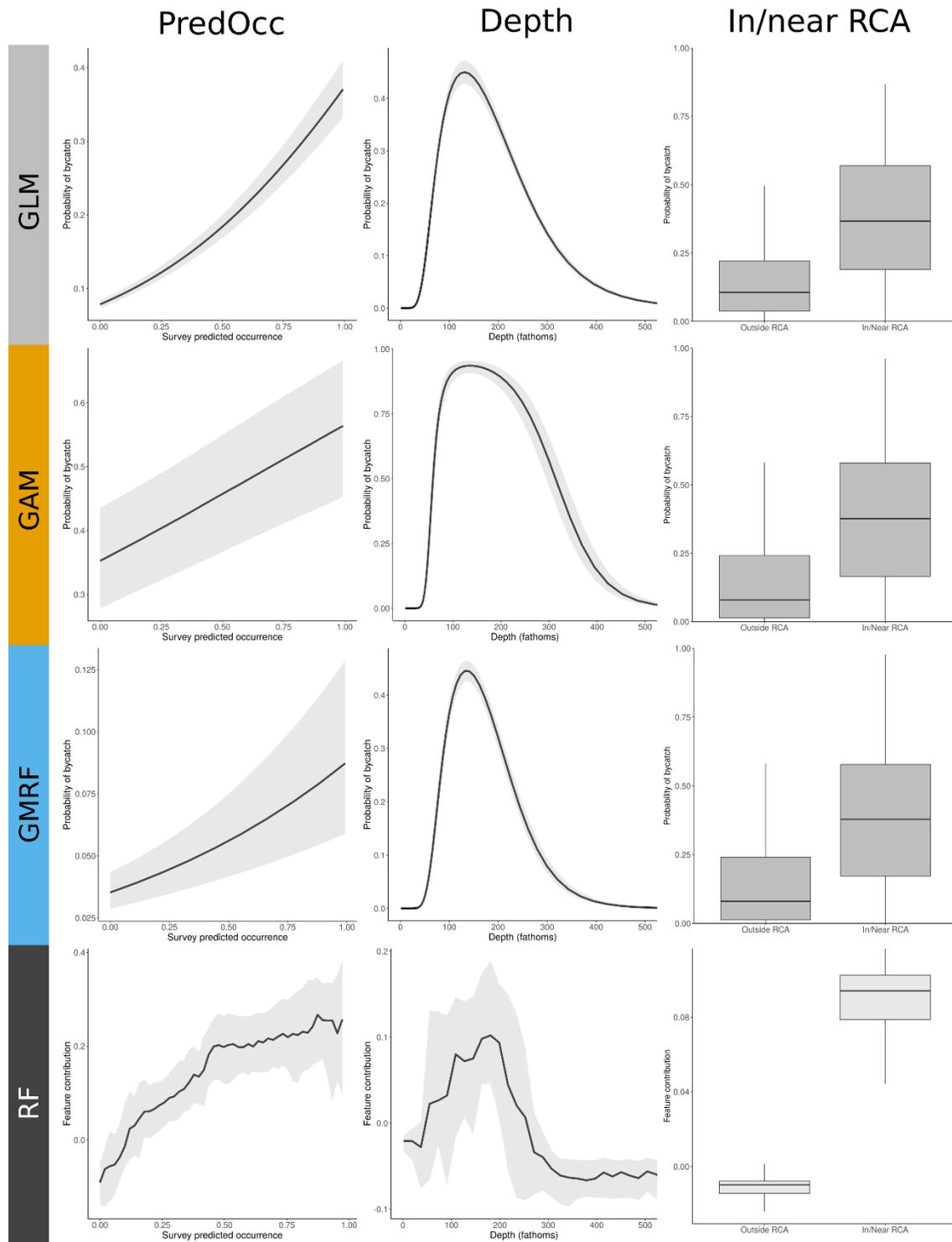
1085

1086
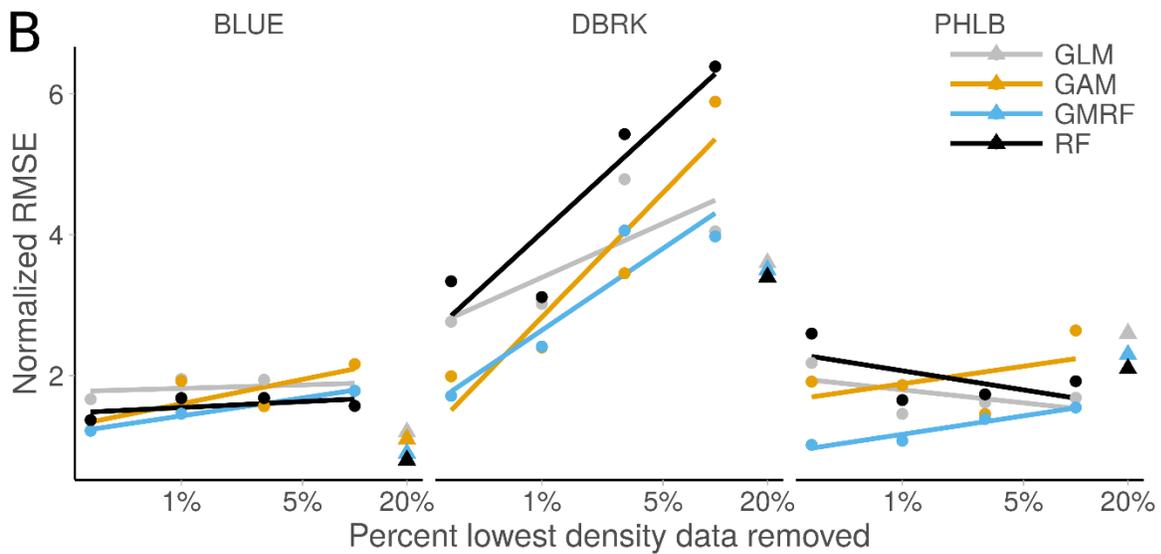
Figure 6.

1089   Figure 7.
1090

1091
1092