



Accounting for spatio-temporal variation and fisher targeting when estimating abundance from multispecies fishery data

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2015-0598.R2
Manuscript Type:	Article
Date Submitted by the Author:	28-Mar-2016
Complete List of Authors:	Thorson, James T.; NOAA, Fisheries Resource Assessment and Monitoring Division Fonner, Robert; Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Conservation Biology Division, Haltuch, Melissa; Northwest Fisheries Science Center, Ono, Kotaro; University of Washington, School of Aquatic and Fishery Sciences Winker, Henning; Department of Statistical Sciences, University of Cape Town, Centre for Statistics in Ecology, Environment and Conservation (SEEC),
Keyword:	Spatio-temporal model, fisher behavior, fishing tactics, index standardization, catch-per-unit-effort

SCHOLARONE™
Manuscripts

1 **Accounting for spatio-temporal variation and fisher targeting when**
2 **estimating abundance from multispecies fishery data**

3
4 James T. Thorson^{1,*}, Robert Fonner², Melissa A. Haltuch¹, Kotaro Ono³, Henning Winker^{4,5}

5
6 ¹ Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National
7 Marine Fisheries Service, NOAA, Seattle, WA, USA

8 ² Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service,
9 NOAA, Seattle, WA, USA

10 ³ School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA98195-
11 5020, USA

12 ⁴ South African National Biodiversity Institute (SANBI), Kirstenbosch Research Centre, Claremont 7735,
13 South Africa

14 ⁵ Centre for Statistics in Ecology, Environment and Conservation (SEEC), Department of Statistical
15 Sciences, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa

16
17 * Corresponding author:

18 Telephone: 206-302-1772

19 Email: James.Thorson@noaa.gov

20 # RF, MH, KO, and HW contributed equally to this contribution and are listed alphabetically

21

22 **Abstract**

23 Estimating trends in abundance from fishery catch rates is one of the oldest endeavors in
24 fisheries science. However, many jurisdictions do not analyze fishery catch rates due to
25 concerns that these data confound changes in fishing behavior (adjustments in fishing location or
26 gear operation) with trends in abundance. In response, we develop a spatial dynamic factor
27 analysis (SDFA) model that decomposes covariation in multispecies catch rates into components
28 representing spatial variation and fishing behavior. SDFA estimates spatio-temporal variation in
29 fish density for multiple species, and accounts for fisher behavior at large spatial scales (i.e.,
30 choice of fishing location) while controlling for fisher behavior at fine spatial scales (e.g., daily
31 timing of fishing activity). We first use a multispecies simulation experiment to show that SDFA
32 decreases bias in abundance indices relative to ignoring spatial adjustments and fishing tactics.
33 We then present results for a case study involving petrale sole in the California current, for
34 which SDFA estimates initially stable and then increasing abundance for the period 1986-2003,
35 in accordance with fishery-independent survey and stock assessment estimates.

36

37 **Keywords:** Spatio-temporal model; fisher behavior; multispecies targeting; métier; fishing
38 tactics; index standardization; catch-per-unit-effort; spatial dynamic factor analysis (SDFA)

39

40 **Introduction**

41 Many developed and developing countries collect detailed information from fishers regarding the
42 location, timing, and catches from fishing operations. This fishing information is typically used
43 to estimate total catch of a given species by a fishery (Somers et al. 2014). Total catch is then
44 often used as a primary input into stock assessment models that are used to estimate population
45 abundance and productivity for informing fisheries management (Methot and Wetzel 2013).
46 Results from stock assessment models are generally used to regulate fishing activities by
47 imposing limits on fishery harvest, the timing or location of fishing, or restrictions on fishing
48 gear.

49 For over sixty years, fisheries scientists have tried to estimate information about trends in
50 population abundance for exploited fishes from data regarding fishery catches and effort
51 (Beverton and Holt 1957; Smith 2007). Fishery information generally includes the catch per
52 fishing operation (e.g., a fishing trip or trawl tow), and in some cases changes in catch per effort
53 between years will be informative about changes in population abundance. For example, the
54 common catch equation:

$$E(C) = QFD$$

55 describes catch the expected value of catch C as the product of catchability Q , nominal fishing
56 effort F , and local density D . If fishing effort is known and catchability is constant, an increase
57 in catch is associated with increasing fish abundance. Therefore, a regression model can be
58 fitted to records of catch data, and the resulting estimates of density can be used as an index of
59 abundance in a stock assessment model. Data regarding fishing characteristics that affect fishing
60 effort or catchability (e.g., vessel identity, the timing of fishing, etc.) can be included as
61 covariates in the regression model (Maunder and Punt 2004). If this is done, variation caused by

62 these covariates is “controlled for” (i.e., filtered out) when estimating the resulting index of
63 abundance. This process is typically referred to as “standardizing” the index of abundance
64 (Beverton and Holt 1957; Maunder and Punt 2004).

65 However, standardizing catch and effort data for a fishery poses many complications.
66 Fishers will often restrict fishing to areas with high fish density and, if the fish become more
67 concentrated over time, this will generally result in elevated catch rates and mask evidence of
68 declining population abundance (Walters and Maguire 1996). Analysts may partially account for
69 this by standardizing fishery data using a spatial index standardization method (Walters 2003;
70 Carruthers et al. 2011). Furthermore, the common catch equation is a simplification of the
71 Cobbs-Douglas production function, given the assumption that fishing effort and density both
72 have a linear effect (Wilberg et al. 2010). Catches may in fact be nonlinearly related to local
73 densities and fishing effort (Harley et al. 2001), although these nonlinear effects are not generally
74 estimable without auxiliary information (Kotwicki et al. 2012). Finally, catchability may vary
75 among fishing operations in ways that are not directly observable, e.g., due to small changes in
76 timing of fishing operations.

77 The catchability of fishing operations may also vary due to a number of technical and
78 behavioral characteristics of fishing (Pelletier and Ferraris 2000; Wilberg et al. 2010; Abbott et
79 al. 2015). Research suggests that fishers can affect the assortment of species caught in a
80 multispecies fishery by modifying the location, timing, and gear characteristics of their fishing
81 activity (Sanchirico et al. 2006; Abbott et al. 2015). Fishermen make discrete operational
82 choices that influence how the catch is distributed across space, time, and species. Before a given
83 fishing season begins, eligible fishers choose whether to participate in the fishery and make
84 vessel capitalization decisions (e.g. repairs, equipment purchases). Once the season starts,

85 fishers must then decide where to fish. The expected profit framework for explaining fisher
86 location decisions (Haynie and Layton 2010) assumes that fishers associate each alternative
87 fishing location with an expected level of profit based on the expected catch (i.e. catch rate and
88 species composition) and expected costs of production (e.g. steaming distance, labor, risk to
89 gear) in that area; and that fisher locational choices are made to maximize expected profits. Upon
90 harvesting fish at their initial location, fishers compare the expected catch to what they are
91 catching. Based on the realization of expectations, fishermen can adjust their operation through
92 coarse or fine scale spatial adjustments, timing adjustments (e.g. soak time, time of day), gear
93 adjustment, or by returning to port. The incentives motivating fisher choices are governed by
94 factors across economic (e.g. input prices, ex-vessel fish prices), environmental (e.g. expected
95 distribution of fish, steaming distances, weather) and regulatory (e.g. spatial closures, catch and
96 effort limits, gear restrictions) dimensions. Understanding how these factors influence fisher
97 behavior in a given fishery can inform the development and interpretation of spatio-temporal
98 models of catch distributions.

99 While fishermen generally prefer to harvest species that yield high profits per unit effort,
100 perfect targeting, or the ability of fisherman to select the species they catch, is often not possible
101 in multispecies fisheries. In other words, these fisheries exhibit joint production technology,
102 where fishing effort always yields more than one species. With joint production, fishermen target
103 species assemblages that maximize their profit rather than targeting a single species¹. At large
104 spatial scales, targeting is accomplished through selection of fishing grounds. At fine spatial
105 scales, targeting is accomplished through implementing fishing tactics (e.g. gear, time of day,
106 bearing, small-scale movements). For clarity, this paper refers to coarse spatial scale fisher

¹ In some cases, the profit maximizing assemblage may be the one with the highest overall density of the most valuable species, though this is not always the case. For example, it may be beneficial for fishermen to avoid a protected species.

107 location choices as “spatial adjustments” and fine spatial scale behavior as “tactics”.

108 Multispecies fisheries with joint production technology are associated with spatial clustering of

109 fishers into groups, called *métiers*², which tend to target similar fishing grounds and employ

110 similar tactics.

111 Information about fishing tactics is increasingly used when standardizing fishery data to

112 estimate an index of abundance (He et al. 1997; Winker et al. 2013). Stephens and MacCall

113 (2004) developed a method that first fits a logistic regression model to the presence or absence of

114 the target species, predicted via catch of non-target species. Fishing operations with a low

115 probability of catching the target species are then excluded in a 2nd-stage model that estimates

116 abundance for the target species. Other authors have applied ordination techniques to

117 multispecies catch rate information in a 1st-stage model (Pelletier and Ferraris 2000). Results

118 from this ordination are then used either as a categorical (Maynou et al. 2003) or continuous

119 (Winker et al. 2013) variable in a 2nd-stage regression model. These 2-stage modelling

120 techniques have several draw-backs including:

- 121 1. An inability to propagate uncertainty from the model defining fishing tactics into the
- 122 subsequent regression model used to estimate abundance;
- 123 2. A lack of transparency regarding model design, as well as an inability to simultaneously
- 124 visualize model residuals from both the first and second-stage models;
- 125 3. Difficulties in determining whether the variable or factor representing fishing tactics is
- 126 orthogonal to variables of interest in the second model or, if not orthogonal, what effect this
- 127 lack of independence has on model performance when estimating abundance.

² Pascoe (2015) defines a *Métier* as a “a fishing activity or strategy defined by area and gear, and therefore associated with a particular catch composition as a result”

128 4. Many subjective decisions must be made, i.e., regarding thresholds for excluding tows
129 (Stephens and MacCall 2004), transforming data when estimating fishing tactics (Deporte et
130 al. 2012; Winker et al. 2013), or excluding observations with zero catches, and the effect of
131 these subjective decisions is difficult to evaluate.

132 Despite these draw-backs, research suggests that accounting for fishing tactics leads to
133 performance improvements when estimating abundance indices from fishery data (Haltuch et al.
134 2013; Winker et al. 2014).

135 In this study, we seek to develop a theoretically interpretable and statistically robust method
136 for simultaneously estimating fishing tactics and relative fish abundance when standardizing
137 fishery-dependent catch rate data. The proposed “spatial dynamic factor analysis” (SDFA)
138 model estimates spatial and temporal variation in abundance for multiple fishes caught in a
139 multispecies fishery, while simultaneously accounting for residual correlations in fishery catch
140 rates caused by fishing behavior. After reviewing the context and underlying motivations
141 driving fisher behavior, we use theoretical arguments and a simulation experiment to show that
142 the SDFA model accounts for fisher behavior occurring at coarse spatial scales while controlling
143 for fishing tactics at fine spatial scales (Table 1). We then provide a case-study example
144 involving the winter fishery for petrale sole (*Eopsetta jordani*) 1986-2003, where the resulting
145 fishery-dependent abundance index shows temporal patterns that are consistent with survey data
146 and stock assessment estimates. This case study identifies that fine-scale targeting in this fishery
147 generates a trade-off between petrale and dover sole (*Microstomus pacificus*), while the fishery
148 can also adjust the spatial distribution of fishing to target locations with high petrale density. We
149 provide an R package, *Spatial_DFA*, for implementing SDFA for other cases

150 (https://github.com/James-Thorson/spatial_DFA), and use release number 1.0.0 in the following
151 analysis.

152 **Methods**

153 **Study goals**

154 We seek to estimate an index that is proportional to abundance for a given species using records
155 of catch and fishing effort for multiple species that are caught within a fishery. To do this, we
156 seek to account for four processes that contribute to patterns of fishery catch in multispecies
157 fisheries:

- 158 1. Spatial variation in density for multiple species – Different species will typically occupy
159 different habitats when averaging over long time periods. The spatial mosaic of occupied
160 habitats arises from spatial variation in habitat quality, dispersal ability, and competitive
161 interactions (Soberón 2007).
- 162 2. Autocorrelated, spatio-temporal variation in density for multiple species – A species’
163 distribution will not generally match its average distribution in any given year, and changes
164 in spatial distribution among years can be driven by changes in biological (e.g. spawning
165 dynamics), physical (e.g. ocean temperature) or human-caused (e.g. fishing) processes.
166 Deviations away from a species’ long-term spatial distribution may persist over several
167 sequential years. For example, if juveniles and adults of a species have different spatial
168 distributions, and the species has high variability in cohort strength, then a strong cohort will
169 likely cause autocorrelation in spatio-temporal variation.
- 170 3. Correlations in spatio-temporal variation among species – Both spatial and spatio-temporal
171 variation will in some cases be correlated among species. These correlations arise whenever
172 species have similar habitat associations (Kissling et al. 2012).

173 4. Correlated catch rates arising from fisher tactics – Fishers using trawl or hook-and-line gears
174 sometimes seek to catch one or more high-value species, but also incidentally catch other
175 species due to the joint production nature of many multi-species fisheries. Research suggests
176 that fishers can modify the assortment of species caught during a given occasion by moving
177 to a different fishing location or by modifying fishing tactics within a given fishing location
178 (Sanchirico et al. 2006; Abbott et al. 2015). The existence of fine-scale fishery tactics
179 implies that there will be multispecies correlations in catch rates even after accounting for
180 spatio-temporal variation in density for each species. Negative residual correlations may also
181 arise due to competition among species for fishing hooks.

182 In the following, we distinguish between spatio-temporal and spatially-stratified models as
183 follows:

- 184 • A spatio-temporal model estimates the magnitude and impact of autocorrelation between
185 nearby sites and nearby time periods. As the number of locations increases, the model
186 estimates a smooth function representing spatial distribution, as well as changes in spatial
187 distribution over time (Thorson et al. 2015a).
- 188 • A spatially-stratified model estimates average density within different pre-defined areas. As
189 the number of strata increases, the model will eventually fail to converge as some strata have
190 no data to inform estimates of density (Thorson and Ward 2013).

191 We seek to estimate the distance over which density is correlated (corresponding to the size of
192 habitat patches), and therefore seek to account for the preceding four processes using a spatio-
193 temporal model.

194 **Model structure**

195 We use a statistical model that accounts for fishing behavior, spatial and spatio-temporal
 196 variation, and correlations among species, called “spatial dynamic factor analysis” (SDFA).
 197 SDFA explains density of P species as a log-linear combination of J factors ($0 \leq J \leq P$), where
 198 each factor represents unobserved spatial variation in density for a species assemblage. The j -th
 199 factor follows an autoregressive spatio-temporal process:

$$\psi_j(s, t + 1) = \rho\psi_j(s, t) + \omega_j(s) + \xi_j(s, t)$$

200 where $\psi_j(s, t)$ is the value of factor j at site s and time t , ρ is the magnitude of autocorrelation,
 201 $\omega_j(s)$ is the average value of factor j at site s , and $\xi_j(s, t)$ represents spatio-temporal variation in
 202 that factor (see Table 2 for a list of all symbols used in this model). Both spatial and spatio-
 203 temporal variations follow a spatially correlated stochastic process:

$$\omega_j \sim GRF(0, R(|\mathbf{H}\Delta s|; \kappa))$$

$$\xi_j(t) \sim GRF(0, R(|\mathbf{H}\Delta s|; \kappa))$$

204 where $GRF(0, R)$ is a zero-mean Gaussian random field with correlation function R (see Thorson
 205 et al. (2015a) for more details regarding Gaussian random fields), R is a Matérn correlation
 206 function given distance Δs between two locations and geometric anisotropy \mathbf{H} , where parameter
 207 κ governs the distance over which covariance drop to 10% of its pointwise value (Thorson et al.
 208 2015d).

209 Log-expected density $\log(d_p(s, t))$ for species p is then a linear combination of factors, and
 210 measured spatio-temporal covariates:

$$\log(d_p(s, t)) = \sum_{j=1}^J A_{p,j} \psi_j(s, t) + \sum_{l=1}^L \gamma_{p,l}(t) x_l(s, t)$$

211 where \mathbf{A} is a P by J matrix representing the impact $A_{p,j}$ of factor j on density for species p
 212 (termed the “loadings matrix”), $x_l(s, t)$ is the l -th measured variable (i.e., an intercept, or water

213 temperature) at site s and time t , and $\gamma_{p,l}(t)$ is the effect of covariate l on density for species p in
 214 year t . Loadings matrix \mathbf{A} governs the covariation in log-density among species, and therefore
 215 ensures that density estimates for each species are informed by catch rates for every other
 216 species. We note that the variance of both spatial ω_j and spatio-temporal $\xi_j(t)$ random effects is
 217 one, such that the variance of spatial $\mathbf{A}\omega_j$ and spatio-temporal $\mathbf{A}\xi_j$ variation is controlled by the
 218 value of \mathbf{A} . Log-expected catch $\log(\lambda_p(i))$ for species p and fishing occasion i is then a linear
 219 combination of local log-density, variation in catchability caused by fishing tactics, and observed
 220 variables associated with variation in catchability (e.g., vessel ID):

$$\log(\lambda_p(i)) = \log(d_p(s_i, t_i)) + \sum_{k=1}^K B_{p,k} \varepsilon_k(i) + \sum_{m=1}^M v_{p,m} y_m(i)$$

221 where $y_m(i)$ is the m -th measured variable affecting catchability for fishing occasion i , $v_{p,m}$ is
 222 the effect of covariate m on catch rates for species p , and \mathbf{B} is a P by K loadings matrix,
 223 representing the impact $B_{p,k}$ on species p of small-scale variables $\varepsilon_k(i)$:

$$\varepsilon_k(i) \sim N(0,1)$$

224 such that $\varepsilon_k(i)$ is one or more latent variables representing variation in fishing tactics, where
 225 $0 \leq K \leq P$. Variable $y_m(i)$ therefore accounts for measured processes affecting catchability
 226 (e.g., fishing vessel size), while \mathbf{B} and $\varepsilon_k(i)$ account for residual, unmeasured variation in
 227 catchability. The component of this model representing multispecies spatio-temporal dynamics
 228 (i.e., $d_p(s, t)$) has previously been explored by Thorson et al. (In press-a). The component
 229 representing fishing tactics (i.e., $B_{p,k} \varepsilon_k(i)$) has not previously been included in SDFa, but
 230 resembles correlations used in other spatio-temporal models (e.g., Kristensen et al. 2014).

231 Finally, we specify a distribution for residual variation in catch rates:

$$\Pr(C = c_p(i)) = \begin{cases} \exp(-v_{p,2}\lambda_p(i)) & \text{if } C = 0 \\ \left(1 - \exp(-v_{p,2}\lambda_p(i))\right) \text{Gamma}(C; v_{p,1}^{-2}, \frac{\lambda_p(i)}{1 - \exp(-v_{p,2}\lambda_p(i))} v_{p,1}^2) & \text{if } C > 0 \end{cases}$$

232 where $\text{Gamma}(C,x,y)$ is the PDF of a gamma distribution with shape x and scale y , $v_{p,1}$ is the
 233 coefficient of variation for positive catches for species p , and $v_{p,2}$ controls the relationship
 234 between the probability of encountering zero individuals and predicted catch for species p , such
 235 that the probability of not encountering a species ($C=0$) is identical to a Poisson distribution with
 236 intensity $v_{p,2}\lambda_p(i)$.

237 We estimate parameters using conventional techniques for mixed-effect models (Thorson and
 238 Minto 2015b). Specifically, fixed effects are estimated by identifying their values that maximize
 239 the likelihood function, which specifies the probability that data arose given those values for
 240 fixed effects when integrating across random effects. We estimate fixed effects using Template
 241 Model Builder (TMB; (Kristensen 2014)) from within the R statistical environment (R Core
 242 Development Team 2013). Parameter estimation using TMB involves the following steps: (1)
 243 we specify a template file that specifies the joint probability, calculated as the product of the
 244 probability that the data arose given specified values for fixed and random effects, and the
 245 probability that random effects arose given values for fixed effects; (2) for input values of fixed
 246 effects, TMB optimizes the joint likelihood with respect to random effects; (3) TMB calculates
 247 the 2nd derivatives of the joint likelihood with respect to random effects (the hessian matrix); (4)
 248 TMB uses the hessian matrix to calculate the Laplace approximation to the marginal likelihood
 249 function, and the gradients of this function with respect to fixed effects; (5) we apply a
 250 conventional nonlinear optimizer, using the marginal likelihood function and its gradients, to

251 identify maximum likelihood estimates (MLE) for fixed effects. TMB then predicts random
 252 effects at the values that maximize the joint likelihood conditional on the MLE for fixed effects,
 253 and uses the hessian of the marginal likelihood with respect to fixed effects and the generalized-
 254 delta method to calculate standard errors for all fixed and random effects (Kass and Steffey
 255 1989). For computational convenience, we use the stochastic partial differential equation
 256 (SPDE) approximation (Lindgren et al. 2011) when calculating the probability of Gaussian
 257 random fields (ω_j and $\xi_j(t)$; see Thorson et al. (2014) for more details). Code for implementing
 258 the model is available in the *Spatial_DFA* R package.

259 **Model interpretation**

260 SDFA can be interpreted as a way to factor the catch equation for species p , $\mathbb{E}(C_p) = Q_p F_p D_p$,

261 such that:

$$\log(D_p) = \log(d_p(s, t)) = \sum_{j=1}^J A_{p,j} \psi_j(s, t) + \sum_{l=1}^L \gamma_{p,l}(t) x_l(s, t)$$

$$\log(F_p) = \sum_{k=1}^K B_{p,k} \varepsilon_k(i)$$

$$\log(Q_p) = \sum_{l=1}^M v_{p,m} y_m(i)$$

262 where F_p is nominal fishing effort after accounting for small-scale fisher targeting, and Q_p is
 263 catchability, and where residual variation follows the zero-inflated Gamma distribution. The
 264 covariance of spatio-temporal variation in log-density among species can then be calculated as:

$$\text{Cov}(\log(D_p)) = \mathbf{A}\mathbf{A}^T$$

265 while the covariance of small-scale variation in fishing effort via fisher tactics is:

$$\text{Cov}(\log(F_p)) = \mathbf{B}\mathbf{B}^T$$

266 It is impossible to estimate spatio-temporal covariation (\mathbf{AA}^T) occurring at a finer spatial scale
 267 than the average minimum distance between sites, so remaining covariation is attributed to
 268 fishing tactics (\mathbf{BB}^T). We refer to distances greater than this minimum distance as “coarse
 269 spatial scales”, and smaller distances as “fine spatial scales”.

270 The model also implies that there is no covariation between density and nominal fishing
 271 effort:

$$\text{Cov}(\log(D_p), \log(F_p)) = 0$$

272 i.e., fishers make a decision about the location of fishing, and then make a statistically
 273 independent decision about fishing tactics at that location. We note that this property is likely
 274 violated in real-world applications, where fishers may make different decisions about fishing
 275 tactics depending upon the location of fishing. However, a simple model estimating a linear
 276 linkage between density and fishing effort, $\text{Cov}(\log(D_p), \log(F_p)) = \alpha$, would result in
 277 complete confounding of the link parameter α . We view it as a strength of the SDFFA model that
 278 it allows explicit consideration of potential mechanistic linkages between density and small-scale
 279 fishing tactics, but leave further exploration as a topic for future research. We hypothesize that
 280 the SDFFA model will perform well at estimating trends in abundance when changes in species
 281 composition arise primarily from the spatial allocation of fishing effort.

282 Additionally, an index $I_p(t)$ of total abundance for each species p in time t can be calculated:

$$I_p(t) = \sum_{s=1}^s a(s) d_p(s, t)$$

283 where $a(s)$ is the area associated with each site. The standard error of this estimate can then be
 284 obtained via conventional asymptotic or likelihood-profile techniques.

285 **Simulation experiment**

286 We first conduct a simulation experiment to explore the performance of the SDFA model.
287 To do so, we generate artificial data sets using a simulation model previously developed by Ono
288 et al. (2015). However, we use additional features that were not previously described in that
289 paper. Specifically, we simulate density for four species within each of 100 cells in a 10 cell by
290 10 cell grid. Dynamics for each species follows logistic population growth (given cell-specific
291 density dependence), with movement among cells controlled by the species' habitat preference
292 and mobility, and where fishing effort (and resulting catches) are governed by expected revenue
293 given densities and prices for each species in each year. A single habitat feature (e.g. depth) is
294 simulated as a Gaussian random field, and this habitat variable affects each species' density
295 according to species-specific habitat preferences. For each year in a 30-year simulation interval,
296 the simulation model then involves the following steps:

- 297 1. Simulate the population dynamics of each species using the Schaefer model (Schaefer 1954),
298 where surplus production in each cell is a function of biomass in that cell (i.e., density
299 dependence is local to each cell).
- 300 2. Calculate the total fishing effort in that year, and then distribute fishing effort to each cell
301 given the expected revenue at the start of the year.
- 302 3. Calculate the corresponding catch in each cell, and record catch for use in subsequent
303 analysis while incorporating additional measurement error.
- 304 4. Decrease population abundance in each cell by the amount of fishery catch.
- 305 5. Redistribute species abundance among cells based on species' movement preferences.

306 Further details are provided in Appendix A and B.

307 We present results for a scenario in which fish prices change over time, spurring changes in
308 species targeting and therefore the spatial distribution of fishing effort. Specifically, fish prices

309 are identical for all four species for years 1-10, but in years 11-20 prices go up for Species 1
310 (price is doubled) and down for Species 3 and 4 species (price is halved), while remaining
311 constant for Species 2. Then, in years 21-30, prices return to their original levels. These changes
312 in price induce changes in the spatial distribution of fishing effort. This simulator is available
313 using the function *Generate_scenario_data* in the SpatialDFA R package.

314 This simulation model generates 30 years of catch and effort data distributed among 100 grid
315 cells. The data frame containing catch and fishing effort in each cell and year is then provided to
316 the estimation model, and estimates of total abundance for each species are compared with the
317 “true” simulated abundance for each species. For each simulated data set, we estimate
318 abundance using four configurations of the S DFA model:

- 319 1. “Default model” – Excluding both spatial variation and small-scale tactics ($\mathbf{A} = \mathbf{B} = \mathbf{0}$, such
320 that $J=K=0$);
- 321 2. “Tactics model” – Estimating only covariation caused by fishing tactics ($\mathbf{A} = \mathbf{0}$ but estimating
322 \mathbf{B} , such that $J=0$ and $K=4$);
- 323 3. “Spatial model” – Estimating only spatial variation ($\mathbf{B} = \mathbf{0}$ but estimating \mathbf{A} , such that $J=4$
324 and $K=0$); or
- 325 4. “Spatial and tactics model” – Estimating both spatial variation and fisher tactics (estimating
326 both \mathbf{A} and \mathbf{B} , such that $J=K=4$).

327 For computational reasons, we do not include model selection for number of factors (J) or tactics
328 variables (K) between zero or four, but instead select among major hypotheses (i.e., including or
329 excluding spatio-temporal variation or fisher tactics). We also do not include any catchability
330 covariates (i.e., $M=0$), and the only density covariate is an intercept for each species and year
331 (i.e., $L=I$, where $x_l(s, t) = 1$ and $\gamma_{p,l}(t)$ involves estimating $P \times T$ intercept parameters). We

332 recommend that future research further explore the impact of model selection on S DFA
 333 performance.

334 We replicate this simulation experiment 100 times to explore the estimation performance of
 335 these four alternative models. We evaluate performance by calculating log-relative error for
 336 each model:

$$LRE_{p,t} = \log \left(\frac{\hat{I}_{p,t} / \sum_{t=1}^T \hat{I}_{p,t}}{I_{p,t} / \sum_{t=1}^T I_{p,t}} \right)$$

337 where $LRE_{p,t}$ is the log-relative error for species p in year t , $\hat{I}_{p,t}$ and $I_{p,t}$ are the estimated and true
 338 abundance, respectively for that species and year, $T=30$ is the total number of years, such that
 339 $\hat{I}_{p,t} / \sum_{t=1}^T \hat{I}_{p,t}$ and $I_{p,t} / \sum_{t=1}^T I_{p,t}$ are the estimated and true indices of abundance after rescaling
 340 each to have a mean of one (i.e., given their treatment as relative indices of abundance).

341 **Case study: petrale sole**

342 We also provide a worked example by using the S DFA model to analyze fishery catch rate data
 343 for the winter (Nov-Feb) fishery for petrale sole in waters off of Oregon and Washington from
 344 1986-2003. We restrict analysis to data for fishery logbook records of catches of four categories
 345 (petrale sole, dover sole, sablefish (*Anoplopoma fimbria*), and thornyheads, where the latter
 346 comprises catch of both shortspine (*Sebastolobus alascanus*) and longspine thornyheads (*S.*
 347 *altivelis*)), and hypothesize that fishers target either petrale or dover sole, while potentially
 348 avoiding sablefish (n=47,652 catch rate records for each of four categories). We again do not
 349 include any catchability covariates (i.e., $M=0$), and include an intercept for log-density for each
 350 species and year (i.e., $L=1$, where $x_l(s, t) = 1$ and $\gamma_{p,l}(t)$ involves estimating 4×18 parameters).
 351 To increase computation speed, we estimate abundance at 400 “knots” (i.e., $S = 400$). To do
 352 this, we apply a k-means clustering algorithm to the location of all samples, and then assume that

353 density for each sample is identical to estimated density for the nearest knot (following Thorson
354 et al. (2015d)). Each knot is associated with a 102.1 km² area on average, so processes occurring
355 at smaller scales are attributed to fishing tactics (“fine scale”) and larger scales are attributed to
356 spatial adjustments (“coarse scale”). We specify full rank for the species-tactics matrix (i.e.,
357 $K=4$), and use the Akaike Information Criterion to select the number of spatial factors (i.e.,
358 selection among $J=\{0,1,2,3,4\}$). We then extract the estimate of total abundance for petrale sole
359 from the selected model.

360 Next, we compare estimates of total abundance for the selected model with estimates of total
361 abundance from the triennial bottom trawl groundfish survey. To do so, we fit a previously
362 developed geostatistical index standardization model (Thorson et al. 2015d) to survey data
363 collected during June-Sept from years 1980, 1983, 1986, 1989, 1992, 1995, 1998, and 2001
364 collected by the Alaska Fisheries Science Center (Weinberg 2002). This index standardization
365 model separately analyzes the proportion of survey tows that encounter petrale and the density of
366 petrale when encountered using a delta-GLMM model design. The seasonal timing of the
367 triennial survey changed between 1992 and 1995, so we include calendar date (standardized to
368 have zero mean and unit variance) as a linear covariate in the sub-models for encounter
369 probability and density given encounters.

370 We also compare the estimates of petrale abundance from SDFa analysis of fishery catch
371 rate data with estimates of petrale abundance from the latest stock assessment in 2013 (Haltuch
372 et al. 2013). However, we modify the assessment to exclude all fishery-dependent CPUE
373 information, so that estimates from the stock assessment are statistically independent of the
374 fishery-dependent catch rate data. This stock assessment synthesizes many other data sources,

375 including the triennial survey, as well as age- and length-composition samples to estimate
376 abundance for petrale sole in the California Current.

377 Finally, we conduct a 10-fold cross-validation experiment to evaluate performance of the
378 SDFA model. Cross-validation involves dividing the entire data set of fishery catch-rates into 10
379 equally-sized partitions. For the first replicate, we exclude data for petrale sole from the 1st
380 partition (called the “testing set”), while fitting to data for partitions 2-10 for petrale and all
381 partitions (1-10) for all other species (called the “training set”). This fitted model is used to
382 calculate the natural logarithm of probability of data for petrale in the testing set. We refer to the
383 log-probability of the testing set, divided by the number of observations in the testing set, as the
384 “predictive score”. This process is then repeated for all 10 partitions, using the four alternative
385 model configurations (default, spatial, tactics, and spatial and tactics models). The model that
386 has the highest predictive score is inferred to be the most parsimonious model for out-of-sample
387 prediction (Gelman et al. 2013).

388 **Results**

389 Examining results from the simulation experiment (Fig. 1) shows that all models are
390 approximately unbiased for Species 2, which has no change in pricing among years. However,
391 the model without spatial variation or tactics (Fig. 1: left column) is highly biased for the species
392 that have changes in tactics (Species 1, 3, and 4). This bias is positive for Species 1 and 3 in
393 years 11-20 and negative for years 1-10 and 21-30, where changes in bias coincide with changes
394 in the spatial allocation of fishing (caused by changes in relative price). This pattern of bias is
395 reversed for Species 4, which has the opposite pattern in prices compared with Species 1 and 3.
396 The model with only tactics (Fig. 1: 2nd column) somewhat decreases bias for Species 1, 3, and
397 4, but is still substantially biased. The models with spatial variation, or both spatial and tactics

398 (Fig. 1: 3rd and 4th columns) are least biased, but still show a small positive bias in early years
399 and negative bias in late years for Species 1, and results show that adding tactics (i.e., 4th
400 column) offers little improvement over the model with just spatial variation (3rd column). A
401 follow-up analysis (Appendix C) shows that models with spatial variation, or both spatial and
402 tactics, have decreased performance when using data that are spatially aggregated prior to
403 analysis.

404 Model selection indicates that the model with four dynamic factors ($J=4$) has the most
405 parsimonious fit for the winter petrale fishery. Mapping density for the fishery-dependent index
406 shows that density is highly patchy, with elevated density in small areas off of northern and
407 southern Oregon, and that the hotspot in southern Oregon shows a particularly large increase in
408 abundance from 1993 onward (Fig. 2). A comparison of estimated abundance with estimates
409 from the triennial survey (Fig. 3) illustrates that both models show a small decrease in abundance
410 from 1986-1992, and then an increase from 1995-2001. However, the magnitude of this increase
411 is significantly smaller for the SDFA model. Comparison with estimates of spawning output
412 from the most recent stock assessment corroborate a decrease in abundance prior to 1993, and an
413 increase after 1993, where the magnitude of increase is smaller than either the fishery or survey
414 indices. We therefore conclude that the fishery-dependent index captures many of the same
415 patterns as survey and assessment estimates. Remaining differences between the assessment
416 estimates and the survey-based index of abundance may arise because our inclusion of calendar
417 date has not entirely controlled for changes in survey timing between 1992 and 1995.

418 Finally, our cross-validation experiment shows that the model with both spatial variation and
419 targeting has substantially better predictive performance than a model without spatial variation or
420 targeting (Fig. 4; mean improvement = 0.149 log-score per datum). The model that incorporates

421 targeting explains nearly half of this difference (i.e., mean improvement = 0.074 log-score),
422 while the model that incorporates spatial variation explains nearly three-quarters of this
423 difference (mean improvement = 0.031 log-score). Therefore, including spatial variation leads to
424 a greater improvement in predictive performance than including targeting. Further exploration
425 of spatio-temporal variation in density among species (i.e., $\text{Corr}(\mathbf{f}(s, t))$) shows that petrale
426 density is essentially uncorrelated with dover sole, and is negatively correlated with both
427 thornyheads and sablefish density (Fig. 5). However, covariation due to fisher targeting (i.e.,
428 $\text{Corr}(\mathbf{f}(s, t))$) causes a negative correlation between petrale and all other species (Fig. 6). This
429 supports our hypothesis that fishers target either petrale or other species in the offshore winter
430 bottom trawl fishery.

431 **Discussion**

432 In this study, we have decomposed the catch equation, $\mathbb{E}(C) = QFD$, for each fishing operation
433 (e.g., trawl set or longline soak) into components representing spatio-temporal variation in fish
434 density (D), small-scale tactics that allocate fishing effort (F), measurement variables affecting
435 catchability (Q), as well as residual variation. This decomposition implies that a spatio-temporal
436 model of fish density can account for spatial variation in density (and therefore control for
437 changes in the spatial allocation of fishing effort) while filtering out covariation in model
438 residuals (as caused by fine-scale fisher tactics). We use spatial dynamic factor analysis (SDFA)
439 to implement this model, and provide a simulation experiment and case-study involving petrale
440 sole to show that SDFA can provide reasonable estimates of abundance trends using multispecies
441 fishery data, despite shifts in fishing behavior caused by variation in prices. In both a simulation
442 experiment and a case study involving the winter petrale sole fishery off the US West Coast,
443 accounting for coarse-scale spatial adjustments was more important (lead to a greater decrease in

444 predictive error) than accounting for fine-scale fishing tactics. Importantly, our decomposition
445 assumes that latent variables representing fine-scale fisher tactics are orthogonal to latent
446 variables representing density (i.e., that fishers make independent decisions about where and
447 how to fish). We therefore hypothesize that S DFA will perform well when either (1) the spatial
448 allocation of fishing effort contributes greater variance in catch rates than fine-scale tactics, or
449 (2) large-scale spatio-temporal variation and fine-scale tactics are statistically independent.
450 However, the relative variance of multispecies catch rates caused by large-scale spatial decisions
451 vs. small-scale tactics remains an open research question (Abbott et al. 2015), and relatively little
452 is known about mechanisms for species targeting in many fisheries. Therefore, we cannot
453 confidently speculate when these two conditions will be met in real-world fisheries.

454 Spatio-temporal methods are increasingly used to estimate fish density in fisheries models
455 (Thorson et al. 2015a). Research suggests that spatio-temporal models are more statistically
456 efficient (e.g., have a lower expected root-mean-squared error) than spatially stratified models
457 (see definitions in section titled “Study goals”) in some cases (Shelton et al. 2014; Thorson et al.
458 2015d). When using spatio-temporal models to predict density, the estimated variance of density
459 predictions is generally greater at locations that are geographically distant from available data
460 than locations that are close to available data (e.g., as seen for northern dusky rockfish in the
461 Gulf of Alaska, Lunsford et al. (2015), D. Hanselman, pers. comm. 2015). We believe that this
462 is an appropriate way to characterize uncertainty when sampling is not distributed evenly
463 throughout space, so we consider this characteristic to be a benefit of index standardization using
464 spatio-temporal models. Spatially stratified models remain more common than spatio-temporal
465 models when analyzing fishery catch-per-unit effort data (Carruthers et al. 2011; Ono et al. 2015;
466 Campbell 2016), but exceptions do exist (Kai et al. In review).

467 S DFA also leverages multispecies information in a way that is relatively new in fisheries and
468 ecology. Joint species distribution models are increasingly proposed as a “Robin Hood”
469 approach (Punt et al. 2011) for estimating species distributions, i.e., a way to borrow information
470 from well-surveyed species to inform estimates of density for infrequently encountered species
471 (Warton et al. 2015; Thorson et al. 2015c; Ovaskainen et al. 2015). However, we know of no
472 previous fisheries study that has developed a statistical model to jointly estimate changes over
473 time in the distribution of multiple species (termed a “joint dynamic species distribution model”,
474 JDSMD). In addition to estimating changes in total abundance, JDSMDs can be used to track
475 changes in population range over time, as well as to estimate the total area of the core population
476 range (Thorson et al. In press-a). Tracking both total (or relative) abundance and core area could
477 potentially improve monitoring for cases where stocks radically shift their density in their core
478 range, as has been implicated in both the collapse and recovery of northern cod (Walters and
479 Maguire 1996). The ability to estimate shifts in species range using fishery-dependent catch
480 rates would also greatly expand the data available to monitor climate impacts on marine fish
481 distribution.

482 Despite these benefits, estimates of abundance using a spatio-temporal approach will likely
483 be biased if (1) large areas of the population domain have no fishery data and are therefore
484 unobserved, and (2) trends in the unobserved areas are systematically different from observed
485 areas (see e.g. Campbell (2016) for a discussion). In this case, the S DFA model developed here
486 will imply the assumption that trends are similar between observed and unobserved areas (with
487 greater predictive variance in unobserved areas), and this is a poor assumption in some cases,
488 e.g., during spatial expansion of a fishing fleet (Walters 2003, Kleiber and Maunder 2008).
489 However, we note that the S DFA model could potentially be modified to account for spatial

490 expansion of a fishery. For example, if some spatial areas are entirely unobserved during early
491 years (e.g., during the development of a fishery), then SDFA can be configured such that local
492 log-densities follow a random-walk process (e.g., that $\rho = 1$ and $\omega_k = \mathbf{0}$). In this case, SDFA
493 estimates of density prior to the first year of fishing will be equal to density in the first year.
494 This specification is similar to the “imputation” method of Carruthers et al. (2010), and
495 simulation testing has shown that it decreases bias when imputing density in unobserved areas
496 relative to simple design-based approaches (Thorson et al. In press-b).

497 As with any new statistical method, we recommend several avenues of future research to
498 follow up on the research presented here:

- 499 1. We envision the biggest improvements will arise from incorporating additional covariates to
500 explain spatial variation in either density or catchability. Including habitat variables has been
501 shown to improve density predictions in spatial models (Shelton et al. 2014), and could also
502 improve predictions in unobserved areas whenever there is a strong link between a habitat
503 variable and population density (e.g., Lehodey et al. 1997). Similarly, including variables
504 affecting catchability might be useful both to distinguish trends in fishing power over time
505 (Robins et al. 1998) or predicting small-scale tactics for individual fishing operations
506 (Bigelow and Maunder 2007). We note in particular that individual quota-based
507 management systems provide a natural experiment for estimating the potential impact of
508 fine-scale tactics. In this case, the quantity of quota remaining for target and non-target
509 species could be included as a catchability covariate, and this would provide contrast to
510 estimate the degree to which fishers can influence the species composition of their catch to
511 avoid choke species (Branch et al. 2006; Abbott et al. 2015).

- 512 2. We also acknowledge that many regions do not collect detailed spatial information regarding
513 fishery catches, and that available spatial records might not be accurate. For example, the
514 shark fishery off Southeast Australia reports fishery catches within relatively coarse strata,
515 and this has limited the potential for spatial analysis of these data (Punt et al. 2000). We
516 therefore recommend that future studies explore the impact of missing, coarse-resolution, or
517 erroneous spatial information on the performance of SDFA. In the limit of no spatial
518 information, SDFA will attribute all covariation in catches to fisher tactics (i.e., **B**). In this
519 case, it will be unable to account for changes in the spatial distribution of fishing effort (as
520 also seen in “Targeting” scenario in our simulation study, or Appendix C). This scenario
521 bears some resemblance to the “direct principle components” (DPC) method of Winker et al.
522 (2013, 2014), and we recommend further simulation comparison of non-spatial models for
523 standardizing fishery catch rate data under different data-generating models (in particular the
524 DPC method). We also envision continued research regarding best practices for when to
525 estimate a nonlinear relationship between an estimated index of abundance and stock
526 biomass (Harley et al. 2001; Wilberg et al. 2010), which will likely be necessary when using
527 a non-spatial model for index standardization.
- 528 3. By default, SDFA estimates that loadings parameters (**A** and **B**) are constant over time.
529 However, future studies could explore either random-walk or time-blocked approaches to
530 changes in spatio-temporal covariation or fisher tactics over time. For example, small-scale
531 fisher targeting likely underwent a gradual change in the petrale fishery following the
532 implementation of individual quotas in 2003/2004. If future studies jointly analyze data from
533 both before and after the introduction of individual quotas, we envision that the covariance of
534 fishery tactics (**B**) would be estimated with a time-block in that year. Similarly, changes in

535 ecosystem dynamics (i.e., the introduction or recovery of a new predator) might lead to
536 changes in spatio-temporal covariation in density (Swain et al. 2015).
537 Despite these several topics for future research, we believe that SDFA offers a valuable addition
538 to the toolbox for analyzing fishery-dependent catch rates. In particular, we suggest that SDFA
539 is more easily interpreted than previous two-stage approaches to incorporating fisher tactics
540 during index standardization (e.g., Stephens and MacCall 2004). Interpretability provides
541 several benefits, including (1) allowing insight about which variables are estimable or
542 confounded when included as density or catchability covariates; (2) permitting the testing of
543 hypotheses and interpretation of estimates regarding the covariance between species in density or
544 fisher tactics; and (3) allowing inspection of model residuals and application of model selection
545 methods.

546 **Acknowledgements**

547 We thank R. Hovel and M. Scheuerell for insights leading to the development of spatial dynamic
548 factor analysis, and as always thank K. Kristensen for developing Template Model Builder,
549 without which this project would not be possible. We also thank M. Bravington and A. Haynie
550 for discussions regarding fisher targeting, and A. Berger, D. Goethel, and P. Lynch for
551 organizing the symposium entitled “Space Oddity: Recent advances incorporating spatial
552 processes in the fishery stock assessment and management interface”, occurring at the 2015
553 national AFS meeting in Portland, OR. K. Ono was partially funded by the Joint Institute for the
554 Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement
555 NA10OAR4320148, Contribution No. 2570. Finally, we thank A. Haynie and two anonymous
556 reviewers for comments on an earlier draft.

557

558 **Bibliography**

- 559 Abbott, J.K., Haynie, A.C., and Reimer, M.N. 2015. Hidden Flexibility: Institutions, Incentives,
560 and the Margins of Selectivity in Fishing. *Land Econ.* **91**(1): 169–195.
561 doi:10.3368/le.91.1.169.
- 562 Beverton, R.J.H., and Holt, S.J. 1957. *On the Dynamics of Exploited Fish Populations*. Chapman
563 & Hall, London.
- 564 Bigelow, K.A., and Maunder, M.N. 2007. Does habitat or depth influence catch rates of pelagic
565 species? *Can. J. Fish. Aquat. Sci.* **64**(11): 1581–1594. doi:10.1139/f07-115.
- 566 Branch, T.A., Hilborn, R., Haynie, A.C., Fay, G., Flynn, L., Griffiths, J., Marshall, K.N.,
567 Randall, J.K., Scheuerell, J.M., Ward, E.J., and others. 2006. Fleet dynamics and
568 fishermen behavior: lessons for fisheries managers. *Can. J. Fish. Aquat. Sci.* **63**(7): 1647–
569 1668.
- 570 Campbell, R.A. 2016. A new spatial framework incorporating uncertain stock and fleet dynamics
571 for estimating fish abundance. *Fish Fish.* **17**(1): 56–77. doi:10.1111/faf.12091.
- 572 Carruthers, T.R., Ahrens, R.N., McAllister, M.K., and Walters, C.J. 2011. Integrating imputation
573 and standardization of catch rate data in the calculation of relative abundance indices.
574 *Fish. Res.* **109**(1): 157–167.
- 575 Carruthers, T.R., McAllister, M.K., and Ahrens, R.N.M. 2010. Simulating spatial dynamics to
576 evaluate methods of deriving abundance indices for tropical tunas. *Can. J. Fish. Aquat.*
577 *Sci.* **67**(9): 1409–1427. doi:10.1139/F10-056.
- 578 Deporte, N., Ulrich, C., Mahévas, S., Demanèche, S., and Bastardie, F. 2012. Regional métier
579 definition: a comparative investigation of statistical methods using a workflow applied to
580 international otter trawl fisheries in the North Sea. *ICES J. Mar. Sci. J. Cons.* **69**(2): 331–
581 342. doi:10.1093/icesjms/fsr197.
- 582 Gelman, A., Hwang, J., and Vehtari, A. 2013. Understanding predictive information criteria for
583 Bayesian models. *Stat. Comput.*: 1–20. doi:10.1007/s11222-013-9416-2.
- 584 Haltuch, M.A., Ono, K., and Valero, J. 2013. Status of the US petrale sole resource in 2012.
585 Northwest Fisheries Science Center, National Marine Fisheries Service, Seattle, WA.
- 586 Harley, S.J., Myers, R.A., and Dunn, A. 2001. Is catch-per-unit-effort proportional to
587 abundance? *Can. J. Fish. Aquat. Sci.* **58**(9): 1760–1772.
- 588 Haynie, A.C., and Layton, D.F. 2010. An expected profit model for monetizing fishing location
589 choices. *J. Environ. Econ. Manag.* **59**(2): 165–176.
- 590 He, X., Bigelow, K.A., and Boggs, C.H. 1997. Cluster analysis of longline sets and fishing
591 strategies within the Hawaii-based fishery. *Fish. Res.* **31**(1–2): 147–158.
592 doi:10.1016/S0165-7836(96)00564-4.
- 593 Kai, M., Thorson, J.T., Piner, K., and Maunder, M.N. In review. Predicting the spatiotemporal
594 distributions of oceanic pelagic sharks in the western and central North Pacific.
- 595 Kass, R.E., and Steffey, D. 1989. Approximate bayesian inference in conditionally independent
596 hierarchical models (parametric empirical bayes models). *J. Am. Stat. Assoc.* **84**(407):
597 717–726. doi:10.2307/2289653.
- 598 Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerney, G.J., Montoya,
599 J.M., Römermann, C., Schiffrs, K., Schurr, F.M., and others. 2012. Towards novel
600 approaches to modelling biotic interactions in multispecies assemblages at large spatial
601 extents. *J. Biogeogr.* **39**(12): 2163–2178.

- 602 Kleiber, P., and Maunder, M.N. 2008. Inherent bias in using aggregate CPUE to characterize
603 abundance of fish species assemblages. *Fish. Res.* **93**(1-2): 140–145.
- 604 Kotwicki, S., De Robertis, A., Ianelli, J.N., Punt, A.E., Horne, J.K., and Jech, J.M. 2012.
605 Combining bottom trawl and acoustic data to model acoustic dead zone correction and
606 bottom trawl efficiency parameters for semipelagic species. *Can. J. Fish. Aquat. Sci.*
607 **70**(2): 208–219.
- 608 Kristensen, K. 2014. TMB: General random effect model builder tool inspired by ADMB.
609 Available from <https://github.com/kaskr/adcomp>.
- 610 Kristensen, K., Thygesen, U.H., Andersen, K.H., and Beyer, J.E. 2014. Estimating spatio-
611 temporal dynamics of size-structured populations. *Can. J. Fish. Aquat. Sci.* **71**(2): 326–
612 336. doi:10.1139/cjfas-2013-0151.
- 613 Lehodey, P., Bertignac, M., Hampton, J., Lewis, A., and Picaut, J. 1997. El Niño Southern
614 Oscillation and tuna in the western Pacific. *Nature* **389**(6652): 715–718.
- 615 Lindgren, F., Rue, H., and Lindström, J. 2011. An explicit link between Gaussian fields and
616 Gaussian Markov random fields: the stochastic partial differential equation approach. *J.*
617 *R. Stat. Soc. Ser. B Stat. Methodol.* **73**(4): 423–498. doi:10.1111/j.1467-
618 9868.2011.00777.x.
- 619 Lunsford, C.R., Hulson, P.-J.F., Shotwell, S.K., and Hanselman, D.H. 2015. 12. Assessment of
620 the Dusky Rockfish stock in the Gulf of Alaska. 2015 Plan Team, Alaska Fisheries
621 Science Center, National Marine Fisheries Service, Seattle, WA. Available from
622 http://www.afsc.noaa.gov/refm/stocks/plan_team/GOAdusky.pdf.
- 623 Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent
624 approaches. *Fish. Res.* **70**(2-3): 141–159. doi:10.1016/j.fishres.2004.08.002.
- 625 Maynou, F., Demestre, M., and Sánchez, P. 2003. Analysis of catch per unit effort by
626 multivariate analysis and generalised linear models for deep-water crustacean fisheries
627 off Barcelona (NW Mediterranean). *Fish. Res.* **65**(1): 257–269.
- 628 Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for
629 fish stock assessment and fishery management. *Fish. Res.* **142**: 86–99.
- 630 Ono, K., Punt, A.E., and Hilborn, R. 2015. Think outside the grids: An objective approach to
631 define spatial strata for catch and effort analysis. *Fish. Res.* **170**: 89–101.
632 doi:10.1016/j.fishres.2015.05.021.
- 633 Ovaskainen, O., Roy, D.B., Fox, R., and Anderson, B.J. 2015. Uncovering hidden spatial
634 structure in species communities with spatially explicit joint species distribution models.
635 *Methods Ecol. Evol.*: n/a–n/a. doi:10.1111/2041-210X.12502.
- 636 Pelletier, D., and Ferraris, J. 2000. A multivariate approach for defining fishing tactics from
637 commercial catch and effort data. *Can. J. Fish. Aquat. Sci.* **57**(1): 51–65.
638 doi:10.1139/f99-176.
- 639 Punt, A.E., Smith, D.C., and Smith, A.D.. 2011. Among-stock comparisons for improving stock
640 assessments of data-poor stocks: the “Robin Hood” approach. *ICES J. Mar. Sci. J. Cons.*
641 **68**(5): 972.
- 642 Punt, A.E., Walker, T.I., Taylor, B.L., and Pribac, F. 2000. Standardization of catch and effort
643 data in a spatially-structured shark fishery. *Fish. Res.* **45**(2): 129–145.
644 doi:10.1016/S0165-7836(99)00106-X.
- 645 R Core Development Team. 2013. R: A Language and Environment for Statistical Computing. R
646 Foundation for Statistical Computing, Vienna, Austria. Available from [http://www.R-](http://www.R-project.org/)
647 [project.org/](http://www.R-project.org/).

- 648 Robins, C.M., Wang, Y.G., and Die, D. 1998. The impact of global positioning systems and
649 plotters on fishing power in the northern prawn fishery, Australia. *Can. J. Fish. Aquat.*
650 *Sci.* **55**(7): 1645–1651.
- 651 Sanchirico, J.N., Holland, D., Quigley, K., and Fina, M. 2006. Catch-quota balancing in
652 multispecies individual fishing quotas. *Mar. Policy* **30**(6): 767–785.
- 653 Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models
654 improve estimates of species abundance and distribution. *Can. J. Fish. Aquat. Sci.* **71**(11):
655 1655–1666. doi:10.1139/cjfas-2013-0508.
- 656 Smith, T.D. 2007. *Scaling Fisheries: The Science of Measuring the Effects of Fishing, 1855-*
657 *1955. In 1st edition. Cambridge University Press, Cambridge, UK.*
- 658 Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol.*
659 *Lett.* **10**(12): 1115–1123.
- 660 Somers, K.A., Bellman, M., Jannot, J., Riley, N., and McVeigh, J. 2014. Estimated discard and
661 catch of groundfish species in the 2013 US west coast fisheries. West Coast Groundfish
662 Observer Program. National Marine Fisheries Service, NWFSC, Seattle, WA.
- 663 Stephens, A., and MacCall, A. 2004. A multispecies approach to subsetting logbook data for
664 purposes of estimating CPUE. *Fish. Res.* **70**(2): 299–310.
- 665 Swain, D.P., Benoît, H.P., and Hammill, M.O. 2015. Spatial distribution of fishes in a Northwest
666 Atlantic ecosystem in relation to risk of predation by a marine mammal. *J. Anim. Ecol.*
667 [accessed 19 November 2015].
- 668 Thorson, J.T., Ianelli, J.N., Larsen, E., Ries, L., Scheuerell, M.D., Szuwalski, C., and Zipkin, E.
669 In press-a. Joint dynamic species distribution models: a tool for community ordination
670 and spatiotemporal monitoring. *Glob. Ecol. Biogeogr.*
- 671 Thorson, J.T., Ianelli, J.N., Munch, S.B., Ono, K., and Spencer, P.D. 2015a. Spatial delay-
672 difference models for estimating spatiotemporal variation in juvenile production and
673 population abundance. *Can. J. Fish. Aquat. Sci.* **72**(12): 1897–1915. doi:10.1139/cjfas-
674 2014-0543.
- 675 Thorson, J.T., and Minto, C. 2015b. Mixed effects: a unifying framework for statistical
676 modelling in fisheries biology. *ICES J. Mar. Sci. J. Cons.* **72**(5): 1245–1256.
677 doi:10.1093/icesjms/fsu213.
- 678 Thorson, J.T., Pinsky, M.L., and Ward, E.J. In press-b. Model-based inference for estimating
679 distribution changes in marine species. *Methods Ecol. Evol.*
- 680 Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J., and Kristensen, K. 2015c.
681 Spatial factor analysis: a new tool for estimating joint species distributions and
682 correlations in species range. *Methods Ecol. Evol.* **6**(6): 627–637. doi:10.1111/2041-
683 210X.12359.
- 684 Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015d. Geostatistical delta-generalized
685 linear mixed models improve precision for estimated abundance indices for West Coast
686 groundfishes. *ICES J. Mar. Sci. J. Cons.* **72**(5): 1297–1310.
687 doi:10.1093/icesjms/fsu243.
- 688 Thorson, J.T., Skaug, H.J., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J.H., and Benante,
689 J.A. 2014. The importance of spatial models for estimating the strength of density
690 dependence. *Ecology* **96**(5): 1202–1212. doi:10.1890/14-0739.1.
- 691 Thorson, J.T., and Ward, E. 2013. Accounting for space-time interactions in index
692 standardization models. *Fish. Res.* **147**: 426–433. doi:10.1016/j.fishres.2013.03.012.

- 693 Walters, C. 2003. Folly and fantasy in the analysis of spatial catch rate data. *Can. J. Fish. Aquat.*
694 *Sci.* **60**(12): 1433–1436.
- 695 Walters, C., and Maguire, J.-J. 1996. Lessons for stock assessment from the northern cod
696 collapse. *Rev. Fish Biol. Fish.* **6**(2): 125–137.
- 697 Warton, D.I., Blanchet, F.G., O’Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., and Hui,
698 F.K. 2015. So Many Variables: Joint Modeling in Community Ecology. *Trends Ecol.*
699 *Evol.* [accessed 20 November 2015].
- 700 Weinberg, K.L. 2002. The 2001 Pacific west coast bottom trawl survey of groundfish resources:
701 Estimates of distribution, abundance, and length and age composition. US Department of
702 Commerce, National Oceanic and Atmospheric Administration, National Marine
703 Fisheries Service, Alaska Fisheries Science Center.
- 704 Wilberg, M.J., Thorson, J.T., Linton, B.C., and Berkson, J. 2010. Incorporating time-varying
705 catchability into population dynamic stock assessment models. *Rev. Fish. Sci.* **18**(1): 7–
706 24.
- 707 Winker, H., Kerwath, S.E., and Attwood, C.G. 2013. Comparison of two approaches to
708 standardize catch-per-unit-effort for targeting behaviour in a multispecies hand-line
709 fishery. *Fish. Res.* **139**: 118–131. doi:10.1016/j.fishres.2012.10.014.
- 710 Winker, H., Kerwath, S.E., and Attwood, C.G. 2014. Proof of concept for a novel procedure to
711 standardize multispecies catch and effort data. *Fish. Res.* **155**: 149–159.
712 doi:10.1016/j.fishres.2014.02.016.
713

714

715 Table 1. A summary of the influence of fisher behavior on catch distributions at large and small
 716 spatial scales

	Mechanisms	Model Treatment
Spatial adjustments (i.e. the allocation of effort among sites)	<ul style="list-style-type: none"> • Initial location choice based on expected profit • Spatio-temporal adjustments in fishing location related to changes in relative ex-vessel prices of species, input costs, and regulations over time • Changes in fishing location due to new information obtained from prior fishing (e.g., avoiding areas with low catch rates) 	$\text{Cov}(D_p) = \mathbf{AA}^T$
Tactics (i.e. residual correlations in catch rates at a given site)	<ul style="list-style-type: none"> • Fine-scale spatial adjustments to seek a more favorable species composition and higher catch rates once catch is observed at initial location • Changes in timing of fishing activity (e.g., daytime, nighttime, crepuscular) • Changes in fishing operations, e.g., bearing and speed • Changes in fishing gear (e.g., bait type, hook type, mesh size) 	$\text{Cov}(F_p) = \mathbf{BB}^T$

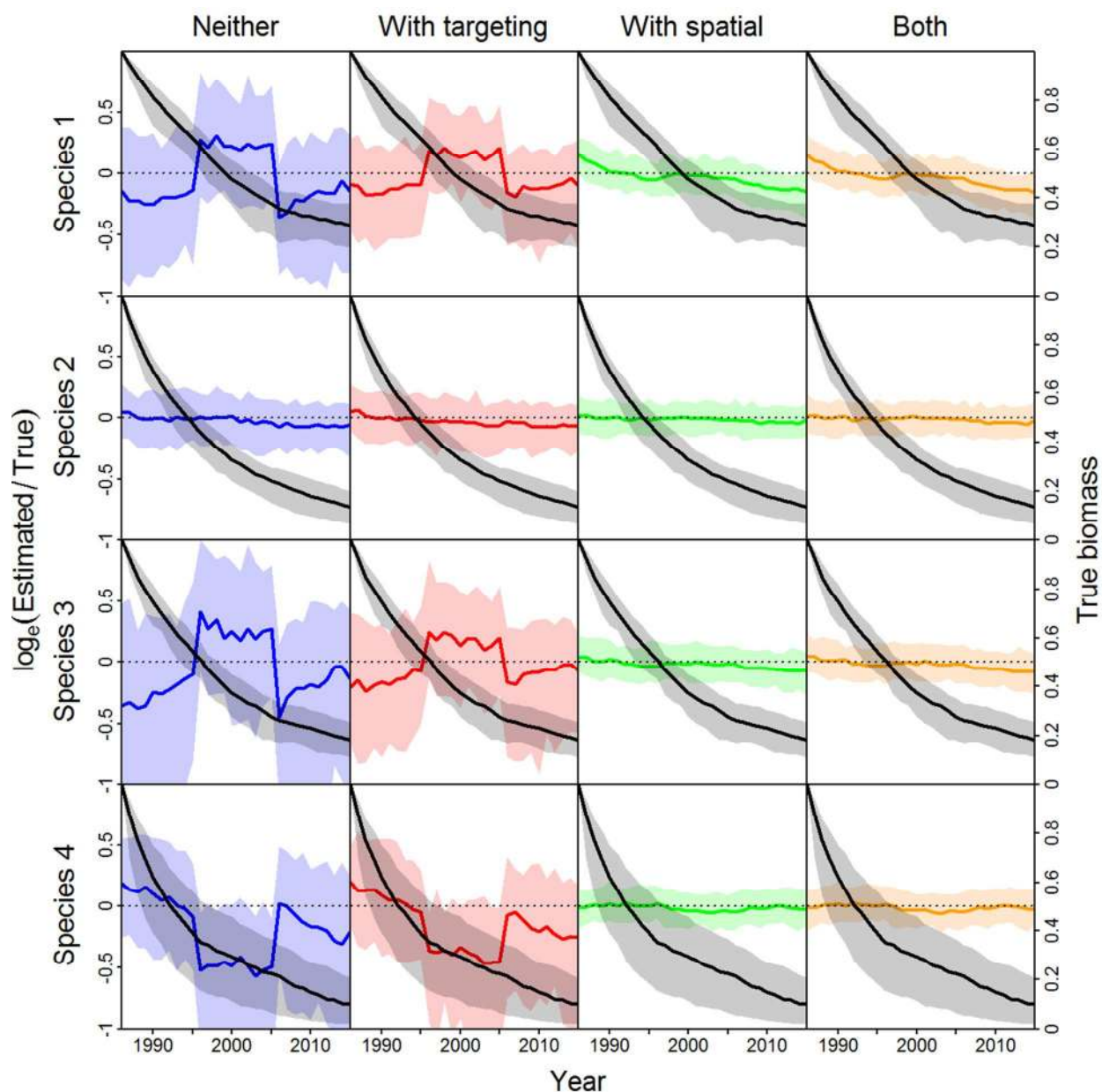
717

718 Table 2 – List of theoretical variables, parameters, data, and indices used in the spatial dynamic
 719 factor analysis model

Name	Symbol
Theoretical variables	
Catch	C
Fishery catchability	Q
Fishing effort	F
Fish density	D
Model parameters	
Spatial dynamic factor	$\psi_j(s, t)$
Spatial variation	$\omega_j(s)$
Spatio-temporal variation	$\xi_j(s, t)$
Temporal autocorrelation for each factor	ρ
Matérn correlation function	R
Geometric anisotropy	\mathbf{H}
Geostatistical range	κ
Log-expected density	$d_p(s, t)$
Loadings matrix for density covariation	\mathbf{A}
Coefficients representing linear effect of density covariates on log-density	$\gamma_{p,l}(t)$
Loadings matrix for targeting covariation	\mathbf{B}
Factor representing targeting	$\varepsilon_k(i)$
Log-expected catch rates	$\lambda_p(i)$
Coefficient of variation for positive catch rates	$v_{p,1}$
Expected rate of decrease of zero catches with increasing density	$v_{p,2}$
Index of abundance	$I_p(t)$
Data	
Density covariate	$x_l(s, t)$
Catchability covariate	$y_m(i)$
Observed catch rates	$c_p(i)$
Area associated with each location	$a(s)$
Dimensions	
Number of spatial dynamic factors	J
Number of locations	S
Number of time intervals	T
Number of species	P
Number of density covariates	L
Number of targeting factors	K
Number of catchability covariates	M
Indices	
Spatial dynamic factor	j
Location	s
Time interval	t
Species number	p
Density covariate	l
Targeting factor	k
Catchability covariate	m

720

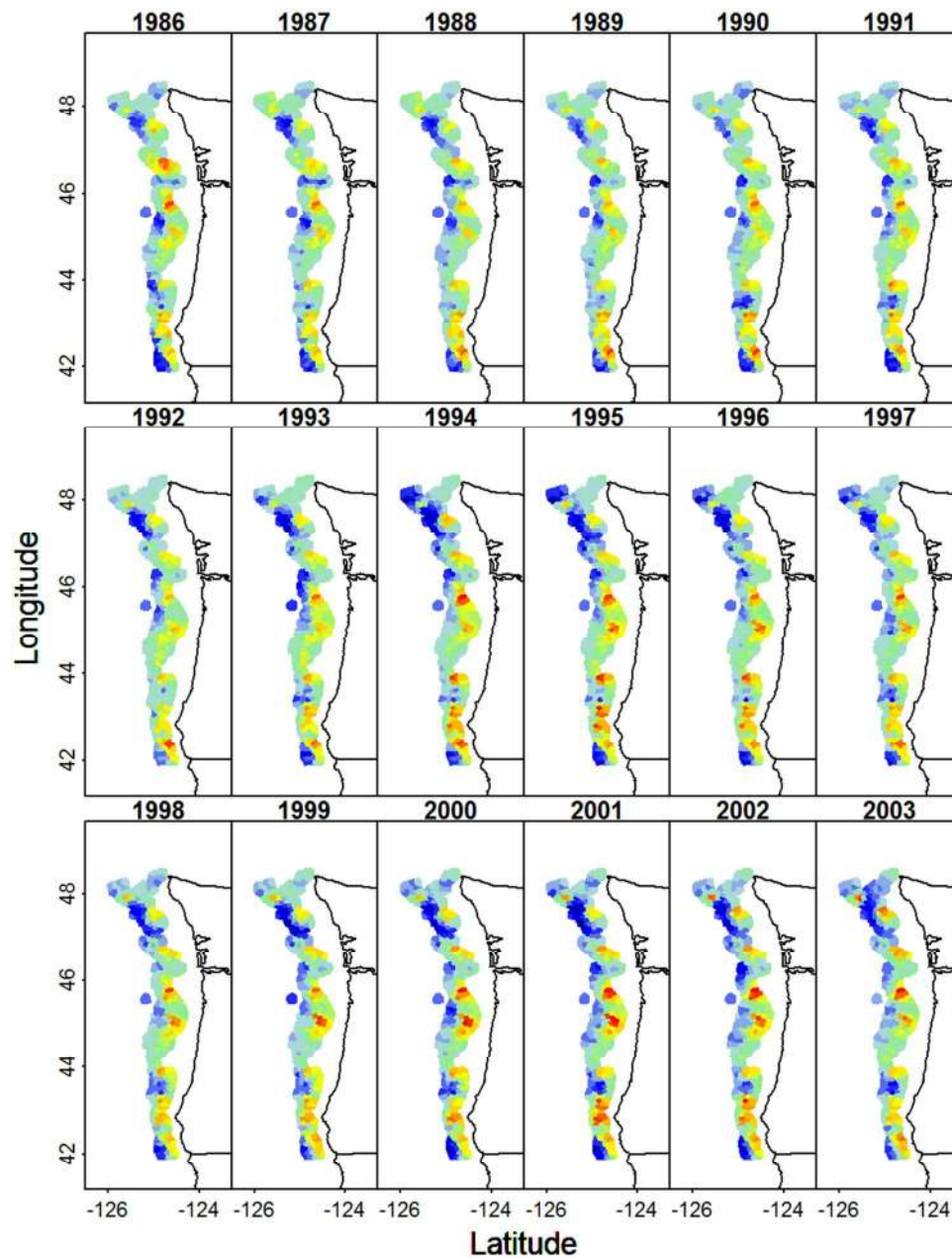
721 Fig. 1 – Simulation experiment for estimating true trends in population abundance (median:
 722 black line; interval containing 90% of simulation replicates: shading; where the biomass trend in
 723 each replicate is rescaled to have a maximum value of one) for four species (rows) using four
 724 configurations of the estimation model (columns) representing no spatial variation or targeting
 725 (1st column), only targeting (2nd column), only spatial variation (3rd column), or both targeting
 726 and spatial variation (4th column). For each estimation model, we rescale the estimated index to
 727 have a mean of one, and then depict the error (defined as the natural log of the rescaled estimate
 728 divided by the rescaled true value) for each year, where a well-performing model will have a line and shaded region near zero (shown as a dotted black line).
 729



730

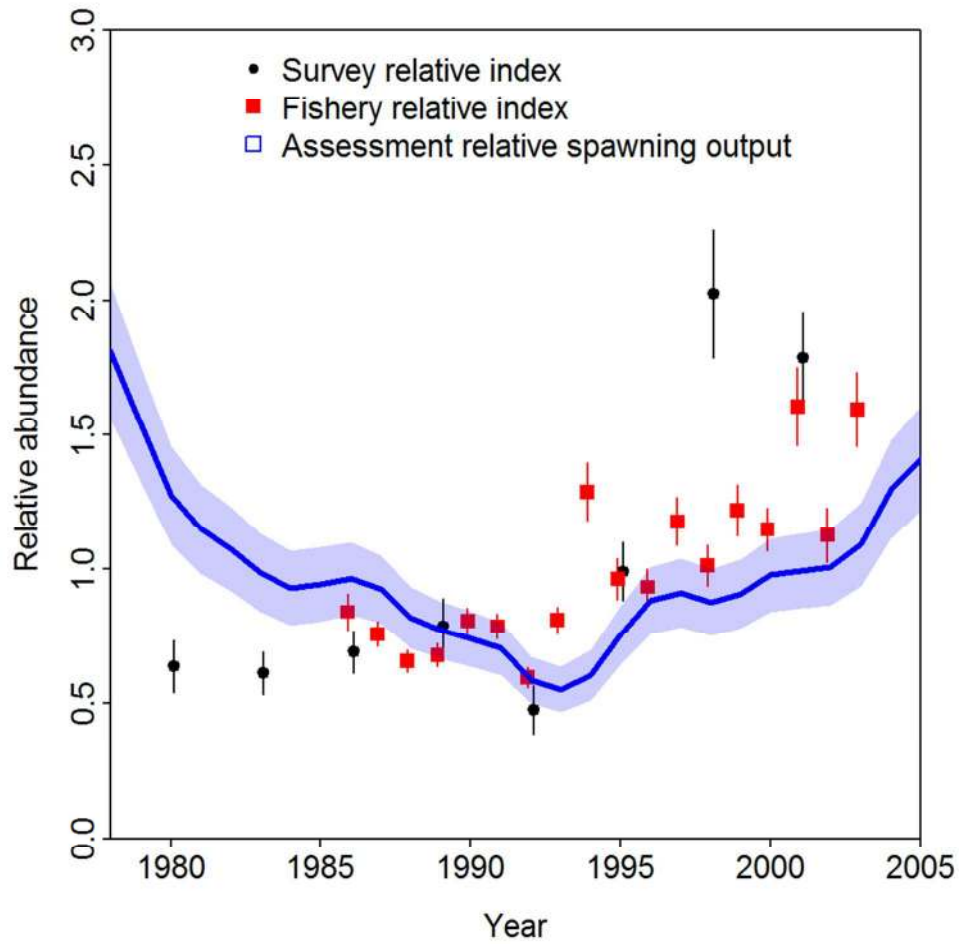
731

732 Fig. 2 – Map of spatio-temporal variation in density for petrale sole, estimated using SDF
733 applied to fishery catch rate data for the winter offshore bottom trawl fishery.



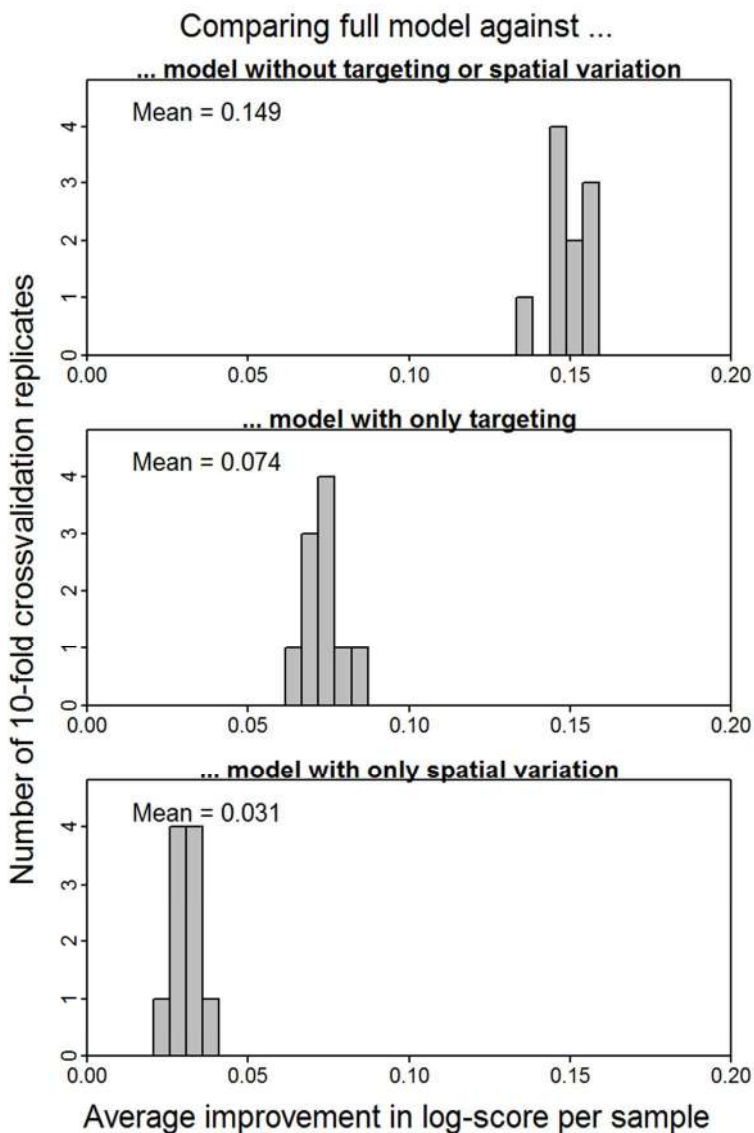
734

735 Fig. 3 – Comparison of survey-based index of abundance, SDFA estimate of fishery-dependent
736 index of abundance, and assessment estimate of spawning output for petrale sole (where each is
737 rescaled to have a mean of one).



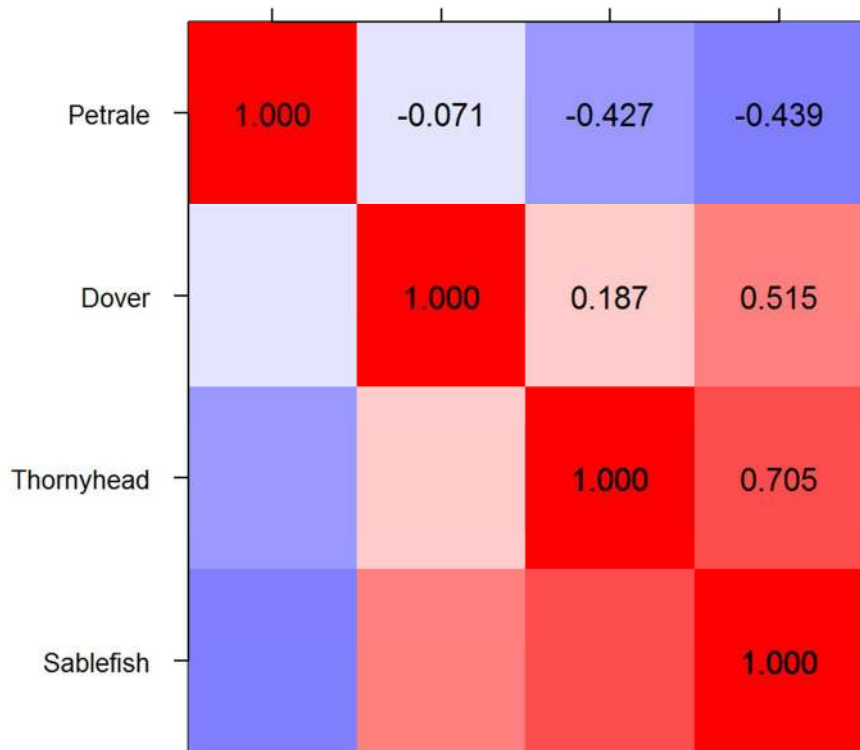
738

739 Fig. 4 – Improvement in log-predictive probability per observation from a 10-fold cross-
 740 validation experiment for a “full” model (which includes multispecies information, i.e.,
 741 estimating \mathbf{A} and \mathbf{B} where $J=K=P$) relative to three alternative models: neither targeting or
 742 spatial variation (top panel: $\mathbf{A} = \mathbf{B} = \mathbf{0}$); only small-scale targeting (middle panel: $\mathbf{A} = \mathbf{0}$,
 743 estimating \mathbf{B}); or only spatial variation (bottom panel: $\mathbf{B} = \mathbf{0}$, estimating \mathbf{A}). Each panel also lists
 744 the average improvement in log-probability per observation across all 10 cross-validation
 745 replicates in the upper-left of each panel.



746

747 Fig. 5 – Spatio-temporal correlation in density among species, $\text{Corr}(\mathbf{d}(s, t))$

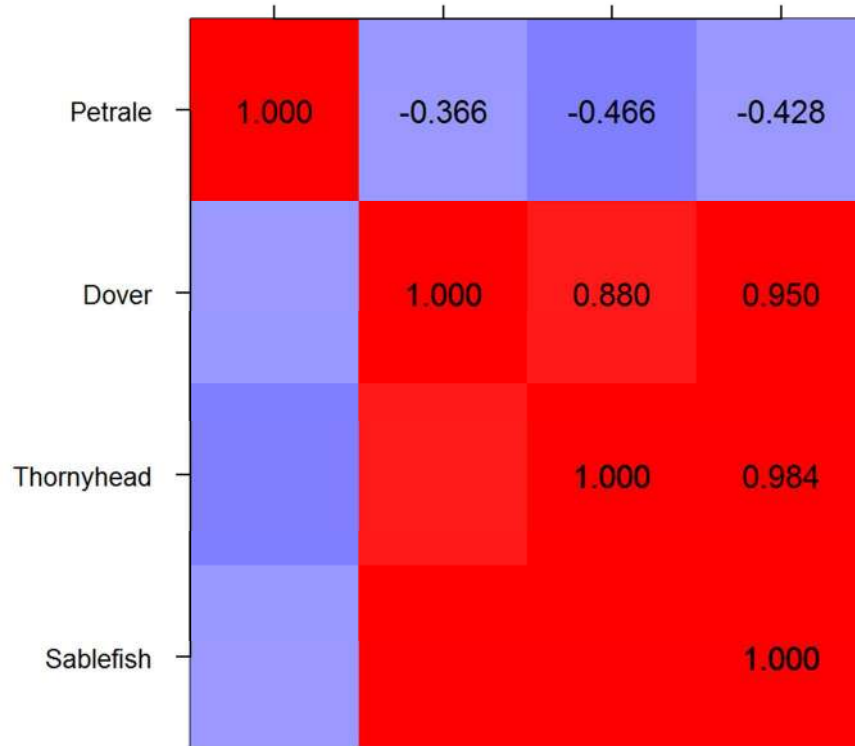


748

749

ait

750 Fig. 6 – Spatio-temporal correlation in fishing effort caused by small-scale targeting among
751 species, $\text{Corr}(\mathbf{f}(s, t))$



752

753

Appendix A: Description of the simulation model

We use a simulation experiment to explore the performance of the proposed spatial dynamic factor analysis model (SDFA) when estimating an index of abundance using multispecies fishery catch rate data. The experiment involves simulating realistic data with known properties using a simulation model, and then fitting SDFA to compare SDFA estimates with known properties from the simulation model. To generate data, we use a simulation model that tracks population abundance for P species in each of S grid cells over T time intervals. Code to use this simulation model is available using function *Generate_scenario_data* in the SpatialDFA R package on the first author's GitHub page (https://github.com/James-Thorson/spatial_DFA), and it is a generalization of the simulator used by Ono et al. (2015) that was developed independently of the SDFA model presented in the main text.

Population dynamics

In this simulation model, the abundance of species p in grid cell s and time t follows a logistic population growth function with movement dynamics controlled by the species-specific depth preference and movement rates. The population density $d_p(s, t)$ for species p in grid s at time t changes through time as a function of the catch $c_p(s, t)$, the maximum per-capita population growth rate r_p , the carrying capacity $k_p(s)$ of species p in cell s , and the proportion of individuals emigrating and immigrating:

$$d_p^*(s, t + 1) = d_p(s, t) + r_p d_p(s, t) \left(1 - \frac{d_p(s, t)}{k_p(s)} \right) - c_p(s, t) - \sum_{s^*=1}^S m_p(s^*, s) d_p(s, t) + \sum_{s^*=1}^S m_p(s, s^*) d_p(s^*, t)$$

The probability $m_p(s^*, s)$ of species p moving from grid s to grid s^* is a function of:

- 20 1. the mobility of the species s , where the probability of moving between cells s and s^* is
 21 modeled as an negative-exponential function of distance between cells and a species-specific
 22 parameter λ_p representing average movement rates;
- 23 2. species depth preferences, where each species has a preferred depth $u_p(z)$, and habitat
 24 preferences decline away from this preferred depth following a lognormal distribution, where
 25 $\sigma_p(z)$ governs the rate at which habitat preferences decline away from preferred depths.
 26 Depth $z(s)$ for each cell s is simulated as a Gaussian random field with an pointwise variance
 27 σ_z^2 , where the covariance between cells follows a negative exponential decay that depends on
 28 distance between grid cells, and where $z(s)$ is added to a constant to ensure that it remains
 29 positive; and
- 30 3. species latitudinal preferences, where each species has a preferred latitude $u_p(x)$, and habitat
 31 preferences decline away from this preferred latitude following a lognormal distribution,
 32 where $\sigma_p(x)$ governs the rate at which habitat preferences decline away from preferred
 33 latitude.

34 Movement rates from cell s^* to cell s is then affected by these three behaviors:

$$m_p(s^*, s) = c_s^{-1} \times \exp(-\lambda_p \Delta s) \times z(s^*)^{-1} \exp\left(\frac{\left(\log(z(s^*)) - \log(u_p(z))\right)^2}{2\sigma_p(z)}\right) \\ \times \exp\left(\frac{\left(x(s^*) - u_p(x)\right)^2}{2\sigma_p(x)}\right)$$

35 where c_s^{-1} is an integration constant defined so that each column of the movement matrix \mathbf{M}_p has
 36 columns that sum to one.

37 Species biomass at the start of the simulation is assumed to be at carrying capacity $k_p(s)$ for
 38 each cell s . Carrying capacity is calculated by determining the stationary distribution of
 39 population density for all cells given productivity and movement parameters.

40 **Effort dynamics**

41 The total number of fishing trips $U(t)$ for all vessels during year t is generated from a discretized
 42 lognormal distribution with mean given by a logistic function of time and a coefficient of
 43 variation of 0.2:

$$U^*(t) \sim \text{Normal}\left(\frac{U_T}{1 + \exp(-0.1t)}, 0.2\right)$$

$$U(t) = \lfloor \exp(U^*(t)) \rfloor$$

44 where U_T is a parameter representing the level of fishing effort in the final year (where we use
 45 $U_T = 100$), and $\lfloor \exp(U^*(t)) \rfloor$ takes the value of $\exp(U^*(t))$ and rounds it down. The vessel
 46 conducting trip u of the $U(t)$ trips in year t is drawn from a uniform categorical distribution:

$$v(u) \sim \text{Categorical}(\boldsymbol{\pi}^*)$$

47 where $\text{Categorical}(\mathbf{x})$ is a categorical distribution where the probability of category i is x_i , and
 48 $\boldsymbol{\pi}^*$ is a vector with length V where each element $\pi_v^* = V^{-1}$. The probability of fishing in cell s in
 49 year t depends on the mean expected revenue in each cell given density $d_p(s, t)$ and price $\phi_p(t)$
 50 of species p in year t :

$$\pi(s, t) = \frac{\sum_{p=1}^P \phi_p(t) d_p(s, t)}{\sum_{s=1}^S \sum_{p=1}^P \phi_p(t) d_p(s, t)}$$

51 Trip u is then assigned to cell s following a categorical distribution:

$$s(u) \sim \text{Categorical}(\boldsymbol{\pi}(t))$$

52 where $\boldsymbol{\pi}(t)$ is a vector with S elements $\pi(s, t)$.

53 Given this allocation of fishing trips among vessels and grid cells, we then simulate fishery
 54 catch for each species and trip. Each vessel has variation in average catchability q_v generated
 55 from a lognormal distribution:

$$\log(q_v) \sim \text{Normal}(\log(0.2), 0.2)$$

56 and catchability $q_p(u)$ for species p on trip u follows a Tweedie distribution:

$$q_p(u) \sim \text{Tweedie}(q_v(u), 1.2, 0.2)$$

57 where $\text{Tweedie}(x, y, z)$ is a Tweedie distribution with expectation x , power y , and dispersion z ,
 58 where this distribution has a variance of $0.1q_v^{1.2}$. Total catch $c_{p,t}(u)$ for trip u of species p in
 59 time t is then calculated:

$$c_{p,t}(u) = d_p(s(u), t)(1 - \exp(-q_p(u)))$$

60 Total catch $c_p(s, t)$ for each species p and cell s and year t is used in the population dynamics
 61 component, and is calculated as:

$$c_p(s, t) = \sum_{u=1}^{U(t)} I(s(u) = s) c_{p,t}(u)$$

62 where $I(s(u) = s)$ is an indicator function that is one if trip u occurs in cell s (i.e., $s(u) = s$) and
 63 zero otherwise. The spatial dynamic factor analysis (SDFA) model then receives a data frame
 64 with a row for each fishing trip and columns representing the year (t), cell $s(u)$, and catch $c_{p,t}(u)$
 65 for each species p .

66 **References:**

67 Ono, K., Punt, A.E., and Hilborn, R. 2015. Think outside the grids: An objective approach to
 68 define spatial strata for catch and effort analysis. *Fish. Res.* **170**: 89–101.
 69 doi:10.1016/j.fishres.2015.05.021.

70

71

72 **Appendix B – Parameter values for each species used in the simulation**73 **experiment**

Name	Symbol	Species 1	Species 2	Species 3	Species 4
Initial price	$\phi_p(1 \leq t < 10)$	1	1	1	1
Middle price	$\phi_p(11 \leq t < 20)$	2	1	0.5	0.5
Final price	$\phi_p(21 \leq t < 30)$	1	1	1	1
Movement rate	λ_p	0.4	0.4	0.4	1.0
Preferred depth	$\mu_p(z)$	220	450	250	1200
Depth range	$\sigma_p(z)$	0.5	0.5	0.3	0.5
Preferred latitude	$\mu_p(x)$	0	0	0	0
Latitudinal range	$\sigma_p(x)$	50	50	50	50
Total initial abundance	$\sum_{s=1}^S d_p(s, 1)$	50,000	820,000	36,000	540,000
Maximum per-capita population growth	r_p	0.22	0.17	0.08	0.08

74

75

76 **Appendix C – Sensitivity of simulation experiment to spatial resolution**

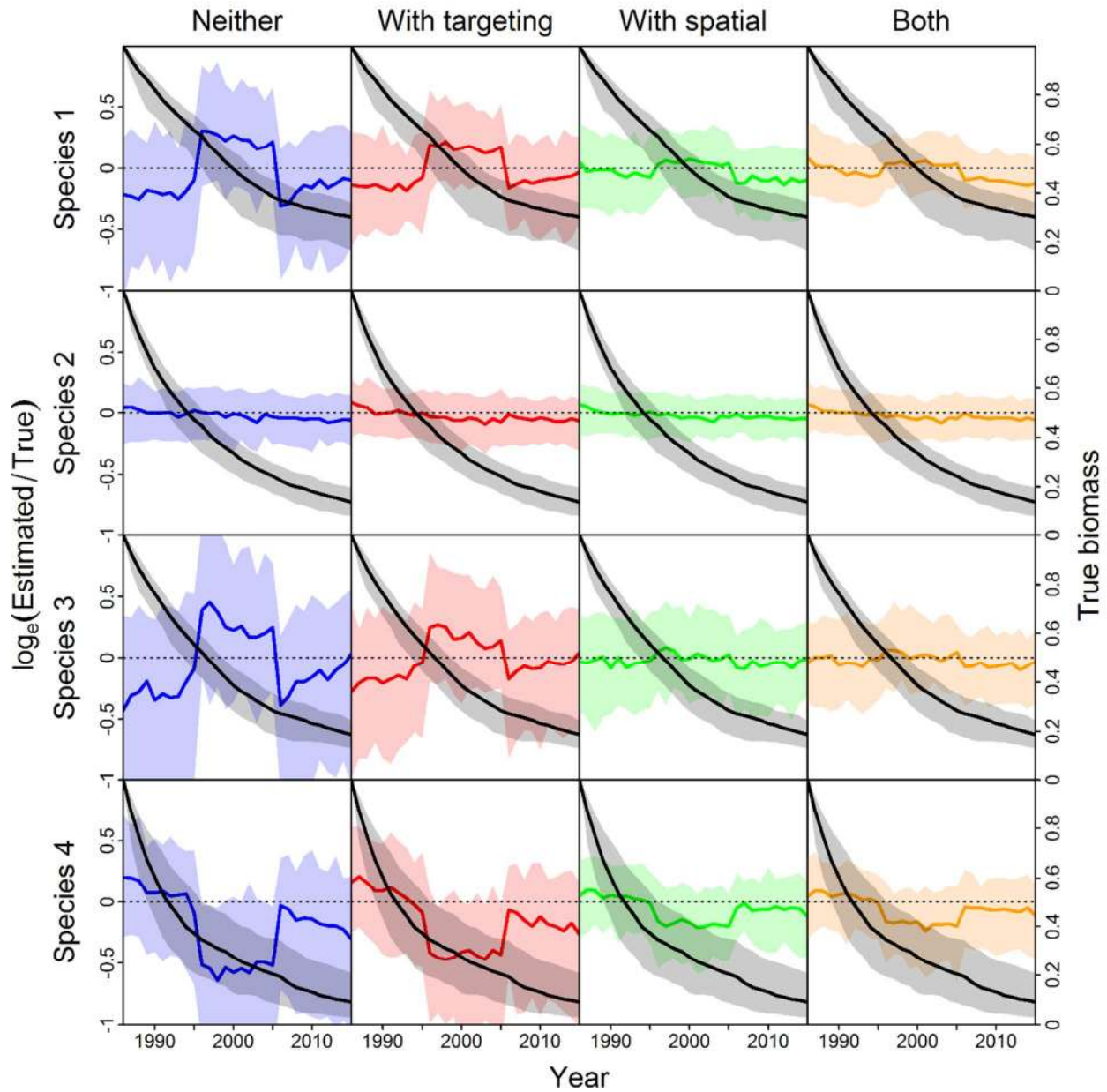
77 We also present results when repeating the simulation experiment but using a coarse resolution
78 for spatial data. Specifically, we simulate data at each of 100 locations on a 10 by 10 grid.
79 However, we then use a k-means algorithm to identify 20 sites, and aggregate each grid cell to
80 the nearest site. In this way, we decrease the spatial resolution of simulated data, so that each
81 site includes data from 5 “sites” on average.

82 This exercise illustrates that decreased spatial resolution for available data has no impact on
83 models that neglect spatial variation (1st and 2nd columns of Fig. S1). However, the performance
84 of the spatial or the spatial-and-targeting models (3rd and 4th columns of Fig. S1) are both
85 decreased. In particular, these spatial models have performance that is more similar to the non-
86 spatial models, with a distinct shift in bias at years 10/11 and 20/21. Biases change at these years
87 due to changes in fish prices, and resulting fisher targeting decisions.

88

89

90 Fig. S1 -- Simulation experiment for estimating true trends in population abundance (see Fig. 1
91 caption for details), given a coarse spatial resolution for spatial variation.



92

93