NOAA Technical Memorandum NMFS-AFSC-434

A Flexible Approach to Optimizing the Gulf of Alaska **Groundfish Bottom Trawl** Survey Design for Abundance Estimation

Z. S. Oyafuso, L.A.K. Barnett, M.C. Siple, and S. Kotwicki

March 2022

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Alaska Fisheries Science Center

The National Marine Fisheries Service's Alaska Fisheries Science Center uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series reflect sound professional work and may be referenced in the formal scientific and technical literature.

The NMFS-AFSC Technical Memorandum series of the Alaska Fisheries Science Center continues the NMFS-F/NWC series established in 1970 by the Northwest Fisheries Center. The NMFS-NWFSC series is currently used by the Northwest Fisheries Science Center.

This document should be cited as follows:

Oyafuso, Z.S., L. A.K. Barnett, M.C. Siple, and S. Kotwicki. 2022. A flexible approach to optimizing the Gulf of Alaska groundfish bottom trawl survey design for abundance estimation. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-434, 142 p.

This document is available online at:

Document available: https://repository.library.noaa.gov

NOAA FISHERIES

A Flexible Approach to Optimizing the Gulf of Alaska **Groundfish Bottom Trawl Survey** Design for Abundance Estimation

Z. S. Oyafuso, L.A.K. Barnett, M.C. Siple, and S. Kotwicki

Resource Assessment and Conservation Engineering Division
Alaska Fisheries Science Center
7600 Sand Point Way N.E.
Seattle, WA 98115

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Alaska Fisheries Science Center

NOAA Technical Memorandum NOAA-TM-AFSC-434

March 2022

ABSTRACT

Fishery-independent surveys are subject to unavoidable fluctuations in sampling effort over time, creating a need for efficient and flexible survey designs. Here we use simulation to evaluate the performance of stratified random survey designs in the Gulf of Alaska (GOA) bottom trawl groundfish survey using alternative stratifications and sampling effort allocations relative to the status quo stratified random sampling design across a range of total sampling effort scenarios (corresponding to 1, 2, or 3 survey boats). In the new approach, we combined a genetic algorithm to optimize the placement of stratum boundaries (defined by depth and longitude) over the simulated data with a multivariate optimal allocation algorithm. This method minimized total survey sample size subject to target precision constraints on abundance indices for a suite of species of high commercial or ecological relevance. Given the proposed and status quo survey designs, performance metrics of bias, precision, and uncertainty of precision were computed across repeated simulations using independent draws with observation error. To determine how the spatial scale of optimization may produce the most precise and accurate abundance estimates at the scale required for informing management decisions, we conducted the optimization at two spatial scales: across the entire GOA and a finer scale within each of five GOA NMFS management areas. In general, newly optimized survey designs at both spatial scales produced abundance estimates with similar precision to the status quo survey, yet also increased the accuracy of abundance estimates and both precision and accuracy of their associated variances by reducing biases present for some species relative to the status quo approach. Overall, the proposed optimal survey effort allocation indicated that higher sampling rates in the western and central GOA and lower sampling rates in southeast GOA helped achieve

precision targets across species. The proposed optimized survey is expected to be practically feasible given that the distance between stations and total expected cruise duration is similar to the status quo. We conclude that the proposed design is expected to improve the accuracy of abundance indices and their variances for many species while requiring similar survey resources to the status quo GOA groundfish bottom trawl survey design.

CONTENTS

INTRODUCTION

Fishery-independent surveys provide the data that form the foundation of the science informing fisheries management. Specifically, these data inform the stock assessment models and ecosystem analyses that guide policy decisions for a diversity of fisheries throughout the United States and the world. Among the largest of these fisheries is Alaska groundfish, which accounted for nearly half of overall U.S. domestic landings and 88% of Alaska's total catch in 2018, generating nearly $1 billion in revenues (Fissel et al. 2019). Maintaining high-quality surveys and establishing evidence-based decision-making tools for responding to changes in survey resources and logistics are major priorities for survey teams internationally (ICES 2020). Abundance indices from surveys are commonly used in stock assessments and fisheries research (Hilborn and Walters 2013); thus, it is important that fisheries surveys are designed and conducted such that the estimated abundance index and associated variance are both accurate and precise across a multitude of species of interest. For example, abundance index variance estimates are commonly used for weighting data in stock assessment models (Francis et al. 2011, Kotwicki and Ono 2019). Further, due to their influence on stock assessments, survey precision and bias are used as quality and performance metrics for comparing survey data products and survey sampling designs (Overholtz et al. 2006, Cao et al. 2014).

Most fishery-independent surveys are multispecies in nature, thus tradeoffs in sampling efficiency among species will affect the total efficacy of the key limiting resource that is survey effort. While groundfishes sampled by demersal surveys share some similarities in that they are found near the seafloor, these assemblages consist of numerous species across disparate families with diverse life histories, population dynamics, spatial distributions, and specific habitat

affinities (e.g., Rooney et al. 2018, Zimmermann 2006, Meuter and Norcross 2002). In addition to the challenge of these inherent tradeoffs among species is the added complication that survey planners may want to prioritize data collection efforts for particular species in a given survey year. For example, if a stock is approaching an overfished state, it may be prudent to modify the survey to increase the expected survey precision for that species. In that case, an objective and explicit means of modifying the survey and evaluating the tradeoffs among the precisions of the lower priority species are needed to guide survey planning.

The Alaska Fisheries Science Center (AFSC) Gulf of Alaska (GOA) bottom trawl survey (BTS) is the primary source of biological data for monitoring over 150 fishes and 360 invertebrates in the region (see von Szalay and Raring 2018 for technical details on existing survey operations). The survey follows a stratified-random survey (STRS) design, consisting of 59 strata defined primarily by depth, bottom terrain (e.g., shelf, gully, slope), and statistical areas originally defined by the International North Pacific Fisheries Commission (INPFC), which were since adopted by NMFS as management areas (Tables 1, 2). Allocations of survey effort across strata are calculated using separate univariate (i.e., single-species) Neyman allocations (Cochran 1977). This allocation of effort was optimized with respect to historical stratum variability, haul costs, stratum area, and total sample size averaged across survey years since 1990 for each species. A weighted mean of the sample allocation was then calculated using the product of the mean biomass of a species and its ex-vessel value as the weighting variable. Historically the GOA BTS typically used three boats (except for 1993 and 2001, which used 4 and 2 boats, respectively), sampling simultaneously over the duration of the summer field season (~ end of May through early August). In 2011 the total effort was reduced to two boats due to budgetary

constraints (von Szalay and Raring 2018) and all surveys since then have typically been carried out with two boats, except for one partial three-boat survey that occurred in 2015 (where the third boat started later and sampled fewer stations than typical). With the change to a two-boat sampling design, strata deeper than 700 m were excluded. These deeper areas accounted for approximately 4% of the total survey area and were allocated 1.7% of the stations in the three-boat design.

Limitations of the existing stratified design restrict its flexibility and quality with respect to producing precise estimates of uncertainty in abundance indices. The strata boundaries were created early in the history of the survey, with little documentation as to how boundaries were delineated. Many strata have been allocated few samples limiting the ability to obtain precise estimation of stratum variances. When total sampling effort was reduced in 2011, deeper strata were removed from the survey because one boat was historically capable of sampling the outer shelf and slope areas (von Szalay et al. 2008). Furthermore, there is limited flexibility to adjust sample allocation and strata boundaries to account for changes in survey resources and management needs in the existing design.

Here, we apply a multivariate optimization approach to multispecies survey design in the GOA BTS to determine whether it can provide an improved framework for abundance estimation that is feasible to implement and flexible to changes in survey effort and management needs. We use a simulation approach to compare the precision and accuracy of abundance estimates across a suite of groundfish between proposed and existing GOA BTS designs. To determine whether the proposed designs are feasible to implement with the same resources as the existing design, we compare the distances traveled among stations between designs. We also

conducted multiple additional survey optimizations to test and compare optimized survey solutions with different stratum variables, spatial scales, and additional sources of uncertainty (observation and estimation error) in the survey optimization data inputs. Finally, proposed future GOA BTS survey designs are recommended, in addition to highlighting opportunities for promising extensions of this optimization framework to other surveys.

MATERIALS AND METHODS

The stratified survey optimization approach was based on Oyafuso et al. (2021) and is first overviewed here before detailing in the subsequent method sections. An operating model (OM) was created to predict population densities in space and time for 26 representative groundfish taxa by fitting univariate vector autoregressive spatiotemporal (VAST) models to survey catch and effort data from the GOA BTS. These density predictions were used as the main data inputs into a stratified survey optimization, which utilized a genetic algorithm to search for candidate stratified survey designs. Depth and longitude were used as stratum variables along which the algorithm searched for optimal strata boundaries. The algorithm also employs the Bethel algorithm (Bethel 1989) to find the optimal allocation of survey effort across strata that minimizes total effort according to prespecified precision constraints (i.e., maximum abundance index CVs) across species. Both algorithms work in concert to iteratively search towards more optimal stratified survey designs. The optimization was tuned to find optimal solutions under three different levels of sampling effort, expressed as the total sample size conferred by one, two, or three boats operating simultaneously. Surveys were then simulated

using these proposed optimized survey designs as well as the existing survey design, and the abundance index and CVs were calculated for each species and survey replicate. Bias, precision, and uncertainty of precision estimates of the abundance estimates were calculated and used as performance metrics to compare these proposed STRS designs to the existing STRS design across the species set. Lastly, to evaluate the logistical feasibility of the proposed survey designs, the expected total duration (based on distances between stations) of these surveys was compared between proposed and existing stratified survey designs. The code used to conduct this analysis is available as a code repository on Z.S. Oyafuso's GitHub page: (https://github.com/zoyafuso-NOAA/Optimal_Allocation_GOA).

Table 1. -- Overview of survey characteristics of the existing and proposed optimized stratified survey designs for the Gulf of Alaska bottom trawl survey.

Characteristic	Existing setting	Proposed setting
Number of strata	59 (54 without strata > 700 m)	10, 15 (gulf-wide optimizations); 3, 5 per area (area-wide optimizations)
Number of species	52-57 species	26 species/species complexes, 15 of which were included in survey optimization
Strata characteristics	Depth zone, habitat (slope/shelf/gully), INPFC reporting areas	Depth, longitude (with contemporary management areas reflected in area-level optimizations)
Pre-specified precision constraints	Not integrated in allocation	Specified *a priori*
Operational assumptions	The deepest stratum in each INPFC reporting area was not sampled for the one boat and two boat-effort scenarios; all stations are considered to be trawlable within the sampled strata for the purposes of simulation	All stations are considered trawlable

Survey Area and Species Included

Before conditioning a spatiotemporal model on the available data from the existing survey, the first step in the analysis is to define the species to include and the boundaries of the spatial domain. Twenty-six groundfish species groups comprising cods, flatfishes, rockfishes, sharks, skates, sculpins, and octopods were included in the analysis. This species set encompasses the target species within the GOA groundfish Fisheries Management Plan (NPFMC 2020) except for Pacific halibut, which is managed separately by the International Pacific Halibut Commission (IPHC). Fifteen of the 26 species were prioritized for inclusion in the survey optimization algorithm (hereafter "design species") based on commercial importance and the dependence of stock assessment models on survey-derived abundance indices. The other seven species groups were excluded from the survey optimization but included when simulating surveys and obtaining indices of abundance (hereafter "non-design species").

The spatial domain consisted of the GOA BTS survey area (continental shelf and slope waters less than 1,000 m deep) within the 200 nm Exclusive Economic Zone of the United States from approximately 132ºW-170ºW. Including the entire range of the GOA survey domain in our optimization framework allows testing the effect of the historical practice of not allocating samples to the deepest (> 700 m) strata in the existing one- and two-boat effort survey designs (Table 2).

Table 2. -- Sample allocation for three total sample size scenarios across the existing strata, defined by depth range, management area [common names with area numbers in parentheses: Shumagin (610), Chirikof (620), Kodiak (630), Yakutat (640), Southeast (650)], and bottom terrain (e.g., banks, shallows, gullies, shelf, slopes).

Depth range (m)	Management area	Stratum label	Stratum name	One-boat allocation	Two-boat allocation	Three-boat allocation
1 - 100	Shumagin	10	Fox Islands	5	9	13
		11	Davidson Bank	15	30	44
		12	Lower Alaska Peninsula	7	14	20
		13	Shumagin Bank	11	22	33
	Chirikof	20	Upper Alaska Peninsula	6	12	18
		21	Semidi Bank	5	10	15
		22	Chirikof Bank	12	23	35
	Kodiak	30	Albatross Shallows	7	13	19
		31	Albatross Banks	14	27	40
		32	Lower Cook Inlet	5	9	14
		33	Kenai Peninsula	5	10	16
		35	Northern Kodiak Shallows	3	6	8
	Yakutat	40	Yakutat Shallows	5	9	13
		41	Middleton Shallows	3	6	9
	Southeast	50	Southeastern Shallows	4	7	10
101 - 200	Shumagin	110	Sanak Gully	3	5	7
		111	Shumagin Outer Shelf	11	22	32
		112	West Shumagin Gully	2	3	4
	Chirikof	120	East Shumagin Gully	7	13	20
		121	Shelikof Edge	10	20	30
		122	Chirikof Outer Shelf	10	20	29
	Kodiak	130	Albatross Gullies	11	21	31
		131	Portlock Flats	12	24	35
		132	Barren Islands	7	14	21
		133	Kenai Flats	6	12	17
		134	Kodiak Outer Shelf	10	19	28
	Yakutat	140	Middleton Shelf	4	7	10
		141	Yakataga Shelf	3	6	8
		142	Yakutat Flats	4	7	11
		143	Fairweather Shelf	5	10	15
	Southeast	150	Baranof-Chichagof Shelf	5	10	15
		151	Prince of Wales Shelf	7	14	21

Table 2. – Continued.

201 - 300	Shumagin	210	Shumagin Slope	7	14	21
	Chirikof	220	Lower Shelikof Gully	5	9	14
		221	Chirikof Slope	4	7	10
	Kodiak	230	Kenai Gullies	5	10	15
		231	Kodiak Slope	3	5	8
		232	Upper Shelikof Gully	2	3	5
	Yakutat	240	Yakutat Gullies	4	7	11
		241	Yakutat Slope	5	9	13
	Southeast	250	Baranof-Chichagof Slope	2	3	5
		251	Prince of Wales Slope/Gullies	5	10	14
301 - 500	Shumagin	310	Shumagin Slope	2	4	6
	Chirikof	320	Chirikof Slope	2	4	6
	Kodiak	330	Kodiak Slope	3	5	8
	Yakutat	340	Yakutat Gullies	2	2	2
		341	Yakutat Slope	3	5	7
	Southeast	350	Southeastern Deep Gullies	2	4	6
		351	Southeastern Slope	2	4	6
501 - 700	Shumagin	410	Shumagin Slope	2	2	3
	Chirikof	420	Chirikof Slope	2	3	4
	Kodiak	430	Kodiak Slope	2	2	3
	Yakutat	440	Yakutat Slope	2	2	2
	Southeast	450	Southeastern Slope	2	2	3
701 - 1000	Shumagin	510	Shumagin Slope	0	0	2
	Chirikof	520	Chirikof Slope	0	0	2
	Kodiak	530	Kodiak Slope	0	0	4
	Yakutat	540	Yakutat Slope	0	0	2
	Southeast	550	Southeastern Slope	0	0	2

Conditioning and Operating Models

We conditioned univariate spatiotemporal distribution models on historical survey catch rate data for each species using a predictive process framework (Banerjee et al. 2008) as implemented by a vector-autoregressive spatiotemporal model using the VAST R Package (v. 3.6.1; Thorson and Barnett 2017, Thorson 2019a). In general, these models relate the observed response to covariates and latent spatial processes. Specifically, we implement a spatiotemporal generalized linear mixed-effects model where random effects describe spatial and spatiotemporal variation (spatial variation that is constant or time-varying, respectively) in population density while temporal variation in the mean density is modeled as a fixed effect of survey year. Spatiotemporal fields were modeled as independent and identically distributed among years. The model estimates two parameters to approximate geometric anisotropy (Cressie and Wikle 2011, Thorson et al. 2016). Continuous spatial and spatiotemporal random fields were approximated using a mesh with 500 "knots" (Rue et al. 2009, Lindgren et al. 2011) as calculated with the INLA R package (Rue et al. 2009), where the value of spatial variables between knot locations was calculated using bilinear interpolation. This spatial resolution of the model was chosen after initial sensitivity testing indicated that this produced a good tradeoff between computational efficiency and consistency with design-based estimators (here defined as the mean density weighted by stratum area and the expansion of this to total biomass, Eqs. 7-9). Catch rate data comprised station-specific catch weight per area swept by the trawl gear (von Szalay and Raring 2018). Consistent sampling data were available from 1996, 1999, and every other year from 2003 to 2019 (11 observed years). The "Poisson-link" reformulation of a

conventional delta model was used to model encounter probability and biomass density (Thorson 2017), where a gamma distribution was specified for modeling biomass density. These models were implemented in the R software environment (v. 4.0.2).

We evaluated models without density covariates and those with covariates for depth and the square of depth, as depth often explains substantial variation in groundfish catch rates (Johnson et al. 2019). When depth was included in a model, it was first log-transformed, then centered and scaled to a standard normal distribution. These depths were derived from high-resolution bathymetry (100 m^2) developed for a new set of species distribution models for the 2022 EFH 5-year Review (Laman et al. in prep). The primary sources for this bathymetry raster were depth soundings from digitized NOAA National Ocean Service smooth sheets from early surveys that used a variety of methods (Zimmermann et al. 2019; Zimmermann and Prescott 2014, 2015). These depth values were used both in the model fit, at locations where samples were present, and for prediction across the entire survey area.

To evaluate the need for the inclusion of density covariates for a particular species, a ten-fold cross-validation procedure was used. Folds were partitioned so that each fold had the same proportion of data from each observed year as reflected in the full sample. The predictions from each of the single species models with the lowest average out-of-sample predictive negative log-likelihood (NLL) were compiled into a set of predictions representing the multispecies OM. Specifically, the population density (y_{git}; in units of kg km^{-2}) of each species or species group g ($g: 1, 2, \ldots, G = 15$ design species; see Table 3 for a list of the species included in the optimization) in grid cell i ($i: 1, 2, \ldots, N$ = 22419 cells) at time t ($t: 1, 2, \ldots, T = 11$ observed survey years) was predicted onto the GOA survey spatial domain at a resolution of two nautical

miles (~ 3.7 km; some prediction grid cells had smaller area due to intersections with survey domain boundaries) for each species and observed year. Appendix A lists the spatial distributions over time as well as other model outputs and diagnostics for each species. These predictions were taken to represent “true” densities, which were used as data inputs to the survey optimization algorithm and to evaluate the performance of simulated surveys given those designs.

Table 3. -- Average out-of-sample predictive error (negative log-likelihood, NLL) resulting from ten-fold cross-validation compared between univariate models that excluded or included depth as a quadratic density covariate. Values in bold indicate the model with the smaller predictive error (best-fit model), which was selected for use in the operating model. Species not included in the optimization ("non-design species") are shaded in grey. Models that did not converge are indicated by NA.

		Average predictive error	
Species name	Common name	Depth excluded	Depth included
Gadus chalcogrammus	walleye pollock	**3261**	3403
Gadus macrocephalus	Pacific cod	2821	**2781**
Atheresthes stomias	arrowtooth flounder	4180	**4114**
Hippoglossoides elassodon	flathead sole	2050	**1947**
Glyptocephalus zachirus	rex sole	2027	**1963**
Lepidopsetta polyxystra	northern rock sole	1110	**1056**
Lepidopsetta bilineata	southern rock sole	1503	**1393**
Microstomus pacificus	Dover sole	1578	**1544**
Hippoglossus stenolepis	Pacific halibut	3166	**3103**
Sebastes alutus	Pacific ocean perch	2438	**2436**
Sebastes melanostictus and Sebastes aleutianus	blackspotted (BS) and rougheye (RE) rockfishes	1044	**956**
Sebastes brevispinis	silvergray rockfish	**425**	455
Sebastes variabilis	dusky rockfish	1057	**998**
Sebastes polyspinis	northern rockfish	1167	**1035**
Sebastolobus alascanus	shortspine thornyhead	1031	**978**
Pleurogrammus monopterygius	Atka mackerel	*1082*	***603***
Beringraja binoculata	big skate	*531*	***525***
Albatrossia pectoralis	giant grenadier	*445*	***426***
Enteroctopus dofleini	giant octopus	*279*	***275***
Sebastes variegatus	harlequin rockfish	*791*	***547***
Beringraja rhina	longnose skate	*929*	***920***
Anoplopoma fimbria	Sablefish	*1815*	***1770***
Cottoidea	Sculpins	*785*	***773***
Sebastes borealis	shortraker rockfish	*493*	*NA*
Squalus suckleyi	Pacific spiny dogfish	*810*	***777***
Sebastes ruberrimus	yelloweye rockfish	*NA*	*138*

Survey Optimization Problem

The goal of the multispecies stratified survey design optimization is to find the strata boundaries and sample allocation across strata that minimizes total sample size subject to *G* prespecified precision constraints ($U_1, U_2, \ldots, U_G$):

$$min \sum_{h=1}^{H} n_h \tag{1}$$

$$s.t.$$

$$CV(\mu_g) < U_g \; \forall g \in \{1, \ldots, G\} \tag{2}$$

$$CV(\mu_g) = \frac{\sqrt{var_{STRS}(\hat{\mu}_g)}}{\mu_g} \tag{3}$$

$$var_{STRS}(\hat{\mu}_g) = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 \frac{\sigma_{hg}^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \tag{4}$$

$$\sigma_{hg}^2 = \frac{1}{TN_h - 1} \sum_{t=1}^{T} \sum_{i=1}^{N_h} \left(y_{git} - \mu_{hg}\right)^2 \tag{5}$$

where n_h and N_h are the sample sizes and number of sampling units in stratum *h* (of *H* total strata), respectively (Eqs. 1-2). By leveraging density predictions provided by the OM, this optimization is specified using population-level statistics. μ_g (Eq. 3) is the population mean of species g averaged over the cells in the spatial domain and over observed years. μ_{hg} (Eq. 5) is

the population mean density estimate of species g averaged over the cells in stratum h and over observed years. $var_{STRS}(\hat{\mu}_g)$ (Eq. 4) is the stratified random sampling variance associated with the estimate of the overall population mean, $\hat{\mu}_g$. Careful consideration is needed for this variance, specifically the stratum variance σ^2_{hg}, defined in Eq. 5. The OM provides predicted densities across all cells and observed years for each species and integrates many sources of variation, including that from time (year-to-year), habitat covariates (depth), and additional spatial and spatiotemporal variation. A common issue in survey design optimization is how to integrate data from previous survey years (Francis 2006), thus the stratum variance in Eq. 5 was modified to incorporate both within-stratum (note the summation range between $i = 1$ to N_h) density variation across space and within-grid cell density variation across years (note the summation range between $t = 1$ and T).

Survey Optimization Algorithm

A genetic algorithm is used to optimize the boundaries of strata via user-defined strata variables, while the Bethel algorithm (Bethel 1989) is used to optimally allocate samples across strata to obtain abundance estimates with precisions at least as high as those prespecified for each species. The optimization is conducted within a modified version of the SamplingStrata R package (Barcaroli 2014) maintained in a public GitHub repository (https://github.com/zoyafuso-NOAA/Optimal_Allocation_GOA).

The genetic algorithm uses evolutionary principles such as fitness-based selection, recombination, and mutation to iteratively search for an optimal stratification and sample

allocation (Barcaroli 2014). It is initialized with 100 candidate solutions consisting of random strata boundary configurations (specified *a priori*) based on two stratification variables, bottom depth (m) and longitude (eastings, km) for a user-defined number of strata (section "*Survey Optimization Explorations"* for specifications). These stratification variables were selected because, in the GOA, gradients across both depth and location (e.g., longitude) have been observed to describe major patterns in demersal species composition (Mueter and Norcross 2002). For each candidate solution, the Bethel algorithm (Bethel 1989) is used to optimize the allocation of the minimum sample size across strata, subject to Equations 1-2. Fitness is defined as the resultant sample size from the Bethel algorithm, with solutions with smaller sample sizes having higher fitness. Elitism occurs by taking the solutions with highest fitness (defined *a priori* to be solutions in the top 10th percentile for smallest sample size) and automatically advancing them to the next iteration of the algorithm. The remaining solutions are selected with probability proportional to their fitness values to "procreate" a new set of candidate solutions by applying a "crossover" of the solutions. Random changes in the stratifications, or "mutations", are then applied at a given rate to these new solutions at each iteration. The mutation rate defines how often random changes to the solutions occur and was tuned to $1/(1+H)$, where H is the total number of strata, based on previous tuning guidelines to reach reasonable convergence times (G. Barcaroli, pers. comm.). The process of procreation occurs until 100 candidate solutions are created for the next iteration of the algorithm. The average fitness of the candidate set of solutions along with the most fit (i.e., smallest sample size) solution improves over iterations, and a large number of iterations are chosen (300 iterations) to ensure that the algorithm has searched for an optimal solution.

The optimization minimizes total sample size for a given CV constraint, but often it is more useful in practice to solve the converse problem, that is, minimize CVs across the species for a given sample size. To solve this problem, the CV constraints must be tuned to match the total survey effort level of interest, which is often a fixed input for a given survey year dependent on how many boats are contracted to conduct the survey. Historically, two or three boats are used in the AFSC GOA groundfish bottom trawl survey, so we conducted the optimization given sampling effort (implemented as the number of sampled stations n) corresponding to one (n = 292), two (n = 550), or three (n = 825) survey boats.

Setting the CV constraints across species to achieve a given level of total effort can be difficult due to the multispecies aspect of the survey design. First, it should be clear that a benefit of stratified random sampling over simple random sampling is that less sampling intensity is needed for a given level of precision (Cochran 1977). Thus, the expected CV under a simple random sample ($CV_{SRS,g}$) for the three desired levels of sampling (i.e., 292, 550, or 825 stations) can serve as a reasonable starting point for setting CV constraints across species, calculating using the variance for a simple random sample (SRS):

$$var_{SRS}(\hat{\mu}_g) = \left(1 - \frac{n}{N}\right)\frac{\sigma_g^2}{n}, \tag{6}$$

where σ_g^2 is calculated similarly to Eq. 5 with $H = 1$. Second, the extent to which stratification improves precision varies by species. The best way to evaluate how well the precision of an abundance estimate can be reduced through a stratified survey for a species is to conduct the survey optimization for each species independently (hereafter “single-species optimizations”). In these optimizations, the CV constraint for a species (expressed in Eq. 2) was tuned to match the

three survey effort levels. By focusing on one species at a time, the procedure can fully optimize the stratified survey design for a single species. This resulting CV constraint ($CV_{SS,g}$; where SS indicates a stratified random design optimized for a single species) represents the most precise but still unbiased CV that can be attained and is used as an absolute lower threshold for the CV constraint. Given the substantial interspecific variation in spatial distributions observed, these levels may not be attainable when conducting a multispecies survey optimization, but they can be used as a lower limit on the CV constraint. Taken together, $CV_{SRS,g}$ and $CV_{SS,g}$ represent a realistic range of precisions for a given species and are helpful when setting CV constraints across species.

The optimization was initialized using the CV constraint corresponding to $CV_{SRS,g}$. Then, the CV constraints across species are reduced by a fixed proportion (5%) relative to their current values and $CV_{SS,g}$ and the optimization was conducted with these new CV constraints; this produces a slightly higher total sample size. This process is repeated until the desired sample size is attained (or adequately approximated). The process of repeatedly optimizing CVs across species ($CV_{MS,g}$; MS: multispecies) for a given level of sampling effort is provided in the following steps (based on an effort scenario with 550 stations):

1) Compute a vector of SRS CVs (CV_{SRS}) for each species.

2) Find the optimal single-species stratified design for each element of the vector V_{SS} .

3) Run the optimization algorithm with a temporary CV vector (CV_{tmp}) set to CV_{SRS} ($CV_{tmp} \leftarrow CV_{SRS}$) and record the resulting optimized sample size n.

4) If $n < 550$ reduce CV by an increment 5% that of the difference between CV_{tmp} and CV_{SS}:

$CV_{tmp} \leftarrow 0.95CV_{tmp} + 0.05CV_{SS}$.

Rerun optimization with the new vector of CVs (CV_{tmp}) and record the resulting optimized sample size *n.*

5) Repeat step 4 until $n = 550$. $CV_{MS} \leftarrow CV_{tmp}$, where CV_{MS} is a vector of CVs optimized to $n = 550$.

Alternative Survey Optimization Scenarios

A suite of alternative survey optimization explorations was conducted to 1) provide a range of solutions for GOA survey planners and 2) determine sensitivities of the survey optimization to evaluate the robustness of proposed designs to many sources of uncertainty. Table 4 summarizes the differences among the 11 scenarios explored:

1) Spatial scale of optimization: The optimization was conducted at two spatial scales to evaluate which produced the best estimates of abundance at the scale needed for management. The two focal scales were 1) across the entire GOA survey domain (hereafter "gulf-wide", scenario A in Table 4), and 2) independently for each management area (hereafter "area-level", scenario B-K in Table 4). For the gulf-wide optimizations, 10 and 15 total strata were specified as separate solution scenarios. Area-level optimizations were conducted separately within each of the five management areas described in the GOA Groundfish Fishery Management Plan (NPFMC 2020). The areas are defined by longitude from west to east as follows:

Western (159ºW-170ºW); Chirikof (154ºW-159ºW); Kodiak (147ºW-154ºW); West Yakutat (140ºW-147ºW); and Southeast (132ºW-140ºW). Three and five strata were set for each area as separate solution scenarios. The number of strata in each optimization scale was selected to represent a range of options that earlier research indicated would produce reasonable survey performance when simulated while retaining fewer total strata than the existing design (Oyafuso et al. 2021). The main differences between the settings of the current stratified survey and the proposed survey optimization are highlighted in Table 1.

CV constraints were defined across species for each area for the area-level optimization. These area-level CVs are not used to produce the final results, as their values were incrementally reduced as described in the section above until the desired total sampling effort was attained. The expected CVs for each species were calculated under SRS for the three sample size scenarios assuming that the number of samples within each management area is proportional to its area. These expected CVs are initialized as the CV constraints. Similar to the gulf-wide optimizations, CV constraints across species and areas were then reduced by a fixed proportion (5%) until the desired total sample size was attained. For the single-species optimizations employed to provide a lower reference point for the plausible value of CV, ten strata were used in the gulf-wide optimizations, and five strata per area were used for the area-level optimizations. Subsequent explorations of solutions were done with the area-level optimizations.

2) Stratum variables **(**scenario C in Table 4**)**: as an additional option for solution types, in these optimizations only depth was included as a stratum variable, as opposed to both depth and longitude as in the base case. The "depth-only" option is explored in a subsequent section as an additional option (scenarios E, H, I, K).
3) One deep stratum **(**scenarios D-E in Table 4**)**: In these optimizations, depth values > 300 m were assumed to be 300 m to force the optimization to create only one deep (> 300 m) strata per management area. This scenario was selected to determine whether constraining the optimization to allow for stratification of the continental shelf would improve the configuration of strata and abundance estimates for the species most targeted by fisheries, of which most are more abundant on the continental shelf than slope. Initial analyses highlighted 300 m as a general cutoff separating shelf and slope strata. Further, prior analyses have demonstrated that 300 m is the approximate depth cutoff at which species diversity begins to decline in the GOA, as shallow-dwelling shelf species become less prevalent and slope-associated species become more prevalent (e.g., Mueter and Norcross 2002).
4) Sensitivity of data inputs: the predicted population densities under maximum likelihood estimation were used as the default data inputs to the optimization. We conducted a sensitivity analysis on the data inputs to explore how robust these optimal survey solutions were to additional sources of uncertainty. Alternative inputs to the optimization included population density values simulated via two methods within VAST using the function *simulate_data*: a) simulating measurement error conditional on the fixed and random effects (scenarios F and H in Table 4) and b)

simulating fixed and random effects from the joint precision matrix (scenarios G and I in Table 4).

5) Maximum depth cutoff (scenarios J-K in Table 4): the existing stratified survey design as recently implemented at the two-boat (550 station) effort level excludes the deepest strata in each management area, effectively restricting the survey footprint to 700 m rather than the original depth limit of 1,000 m. We conducted a series of optimizations and simulated surveys excluding sampling units > 700 m from the OM that was fitted to the full historical dataset. The purpose of this exploration is to evaluate whether switching to a new STRS design would disrupt the continuity of the existing two-boat survey design. We calculated ratios of the estimated abundances from both designs to evaluate the relative difference, for potential use as a catchability correction factor in stock assessments, in addition to the standard performance metrics. Scenarios L and M utilize the existing STRS survey design cells up to 1,000 m and 700 m, respectively.
6) Interspecific Tradeoffs of Species-Specific CV Constraints: CV constraints for some species may be much lower than the minimum acceptable precision needed for a downstream analysis like a stock assessment. Thus, a separate optimization was proposed to allocate survey resources to focus on achieving lower CV estimates for more variable species. CV constraints were not incrementally reduced once they reached some acceptable CV ($CV_{acceptable}$) For example, if the $CV_{acceptable} = 0.1$ and the expected CV (from Eq. 3) for arrowtooth flounder was 0.06, its CV constraint was changed to $CV_{acceptable}$ to avoid solutions that may inflate CVs for

other species by over-weighting arrowtooth flounder in the sample allocation. For this exploration, the Bethel algorithm was used to calculate optimal allocation of sampling effort under the area-level, 5 strata per area, two boat-effort solution (scenario B in Table 4). Three levels of $CV_{acceptable}$ were evaluated: 0.10, 0.20, and 0.25.

Table 4. -- Summary of the differences in settings used for each alternative explored. Note: Scenarios A-K refer to the proposed STRS designs, and L and M refer to the existing STRS design.

Survey exploration	Scenario option	Scenario label												
		A	B	C	D	E	F	G	H	I	J	K	L	M
Spatial scale of optimization	gulf-wide	X												
	area-level		X	X	X	X	X	X	X	X	X	X	X	X
Stratum variables	depth and longitude	X	X		X		X	X			X			
	depth only			X		X			X	X		X		
	existing STRS design												X	X
Sensitivity of data inputs	MLE	X	X	X	X	X					X	X	X	X
	measurement error simulated						X		X					
	fixed and random effects simulated							X		X				
One deep stratum: cells deeper than 300 m are set to 300 m	no	X	X	X			X	X	X	X	X	X	X	X
	yes				X	X								
Maximum depth cutoff	1,000 m	X	X	X	X	X	X	X	X	X			X	
	700 m										X	X		X

Survey Simulations and Performance Metrics

Stratified random surveys were replicated 1,000 times under the proposed survey designs (both gulf-wide and area-level) and the existing survey designs under the different strata number scenarios and sample sizes. Simulated densities (r instead of y) included the observation error estimated by VAST and were conditional on the fixed (year and depth-covariate effects) and random effects (spatial and spatiotemporal random fields). For simplicity, all cells were deemed to be available to the survey (i.e., trawlable), although in reality there is a substantial amount of untrawlable habitat that is inaccessible to the standard AFSC bottom trawling gear (Baker et al. 2019, von Szalay and Somerton 2017). The STRS estimated total abundance ($\hat{\tau}_{dgt}$) and associated unbiased estimator for the variance was calculated for each survey replicate d:

$$\hat{\tau}_{dgt} = \sum_{h=1}^{H} N_h \bar{y}_{dhgt} \tag{7}$$

$$\bar{y}_{dhgt} = \frac{1}{n_h} \sum_{i=1}^{n_h} r_{hgit} \tag{8}$$

$$\widehat{var}(\hat{\tau}_{dgt}) = \sum_{h=1}^{H} N_h (N_h - n_h) \frac{s_h^2}{n_h}, \tag{9}$$

where s_h^2 is the sample variance of stratum h. The "true" CV, $CV_{TRUE}(\hat{\tau}_{gt})$, describes the precision of the total abundance estimate of species g at time t across survey replicates, defined

as the standard deviation of the simulated survey estimates relative to τ_{gt}, the true mean density of species *g* at year *t*:

$$CV_{TRUE}\left(\hat{\tau}_{gt}\right) = \frac{\sqrt{(D-1)^{-1}\sum_{d=1}^{D}\left(\hat{\tau}_{dgt}-\overline{\hat{\tau}_{\cdot gt}}\right)^2}}{\tau_{gt}}, \quad (10)$$

where $\overline{\hat{\tau}_{\cdot gt}}$ is the abundance estimate of species *g* at year *t* averaged across survey replicates. Relative root mean square error of CV, $RRMSE(CV(\hat{\tau}_{gt}))$, is a measure of uncertainty of the replicate sample CVs of species *g* at time *t* and is a composite measure of the dispersion and bias of the replicate sample CVs about the true CV:

$$RRMSE\left(CV\left(\hat{\tau}_{gt}\right)\right) = \frac{\sqrt{D^{-1}\sum_{d=1}^{D}\left(CV\left(\hat{\tau}_{dgt}\right)-CV_{TRUE}\left(\hat{\tau}_{gt}\right)\right)^2}}{D^{-1}\sum_{d=1}^{D}CV\left(\hat{\tau}_{dgt}\right)}. \quad (11)$$

Lastly, relative biases (RB) of the mean density and CV estimates relative to their respective true values were calculated as

$$RB\left(\hat{\tau}_{dgt}\right) = 100\%\ \frac{\left(\hat{\tau}_{dgt}-\tau_{gt}\right)}{\tau_{gt}}. \quad (12)$$

Distance Traveled: Existing Versus Proposed Survey Designs

To assess the practical feasibility of the proposed survey designs, one metric we considered was the total cumulative distance traveled to complete each station. Historically, *in situ* station selection decisions have been made based on location, timing (i.e., the number of

stations a boat can finish in a day), weather, and vessel capabilities (e.g., length of line for trawls). In general, boats sample stations from west to east; in some years, they are split up by depth in the area around Kodiak. Since we cannot replicate the station decision choices made for simulated surveys under the proposed survey design, a couple of simplifications were made to allocate stations to boats and the order of stations. First, only the area-level, three strata per area, two-boat proposed survey design was evaluated. Second, stations were assigned randomly to boats, with each boat visiting half of the total stations in the survey. Station order was determined by solving the Traveling Salesperson Problem (TSP) using the TSP R package (Hahsler and Hornik 2020) for the set of stations, using the nearest insertion method. This approach iteratively solves for the shortest possible total distance traveled, prioritizing the closest unsampled station to the previously sampled one at each step (Rosenkrantz et al. 1977). Since we do not require that the "traveling salesperson" return to the first station, we modified the problem to calculate the shortest Hamiltonian path, similar to a boat starting at its most western station and ending in the southeast portion of the GOA.

Three types of distances were calculated. First, for each survey year, the actual total path distance was calculated by summing the distances between the observed order of station locations, ignoring port calls and assuming a direct, linear travel path between stations. Second, a survey from the proposed design (area-level optimization, three strata per area, two boats) was simulated 1,000 times, and the minimal path distance from solving the TSP for each boat was calculated for each replicate survey. To compare the historical total path distances to the optimal survey path distances, a third calculation was made by solving the TSP for the station set historically assigned to each boat for each observed survey year. The historical survey path

distances (calculation 1) will be systematically different from the TSP-solved survey path distances for the same station assignments (calculation 3) due to differences in the decision-making process of station choice but are useful for interpretation of the results.

RESULTS

Conditioning and Operating Models

Gulf of Alaska groundfishes comprised a diverse range of spatial distributions. This results in interspecific tradeoffs in the formation of strata boundaries and allocation of samples among strata. We summarize these distributions broadly below, with more detailed interpretations and figures in Appendix A. Many gadids, flatfishes, and skates were broadly geographically distributed and most prevalent along the continental shelf. In contrast, many of the rockfishes and other taxa had more restricted spatial distributions (e.g., northern rockfish in the western region and silvergray rockfish and Pacific spiny dogfish in the central and eastern region) or were more prevalent in deeper waters along the outer shelf and the continental slope (e.g., shortraker rockfish, sablefish, and shortspine thornyhead, which had little temporal variability in spatial distribution over time). Visual evaluations of the probability integral transform residuals (Hartig 2020) showed no aberrant departures from model assumptions (Appendix A).

For 87% of the species selected for optimization (design species) and 100% of the non-design species for which a comparison was possible, the cross-validation procedure favored the inclusion of depth as a population density covariate in the model as it reduced the out-of-sample

predictive error (Table 3). The species for which the model without the depth covariate was favored included silvergray rockfish and the semi-pelagic walleye pollock. As expected, prediction errors for species with patchier distributions such as Atka mackerel and many rockfishes were somewhat higher than the flatfishes in the species set.

Design-based and model-based abundance estimates based on the historical survey data were largely similar. Model-based abundance CVs were consistently lower than design-based estimators and tended to show lower variability over the time series (Fig. 1). Design-based estimators for some of the rockfish species tended to have slightly higher abundance estimates than the model-based estimates, but note that the scale of an index can be influenced by multiple factors, including the spatial resolution of the model and the choice of likelihood function (Thorson et al. 2021). For some of the rockfishes, along with spiny dogfish and Atka mackerel, years with anomalously high design-based estimates of abundance were often associated with lower sample precisions. Overall, the similarity of temporal trends between design-based and model-based abundance estimates, in concert with spatial residual diagnostics (Appendix A), supports the use of these models as OMs in subsequent optimization and simulation procedures.

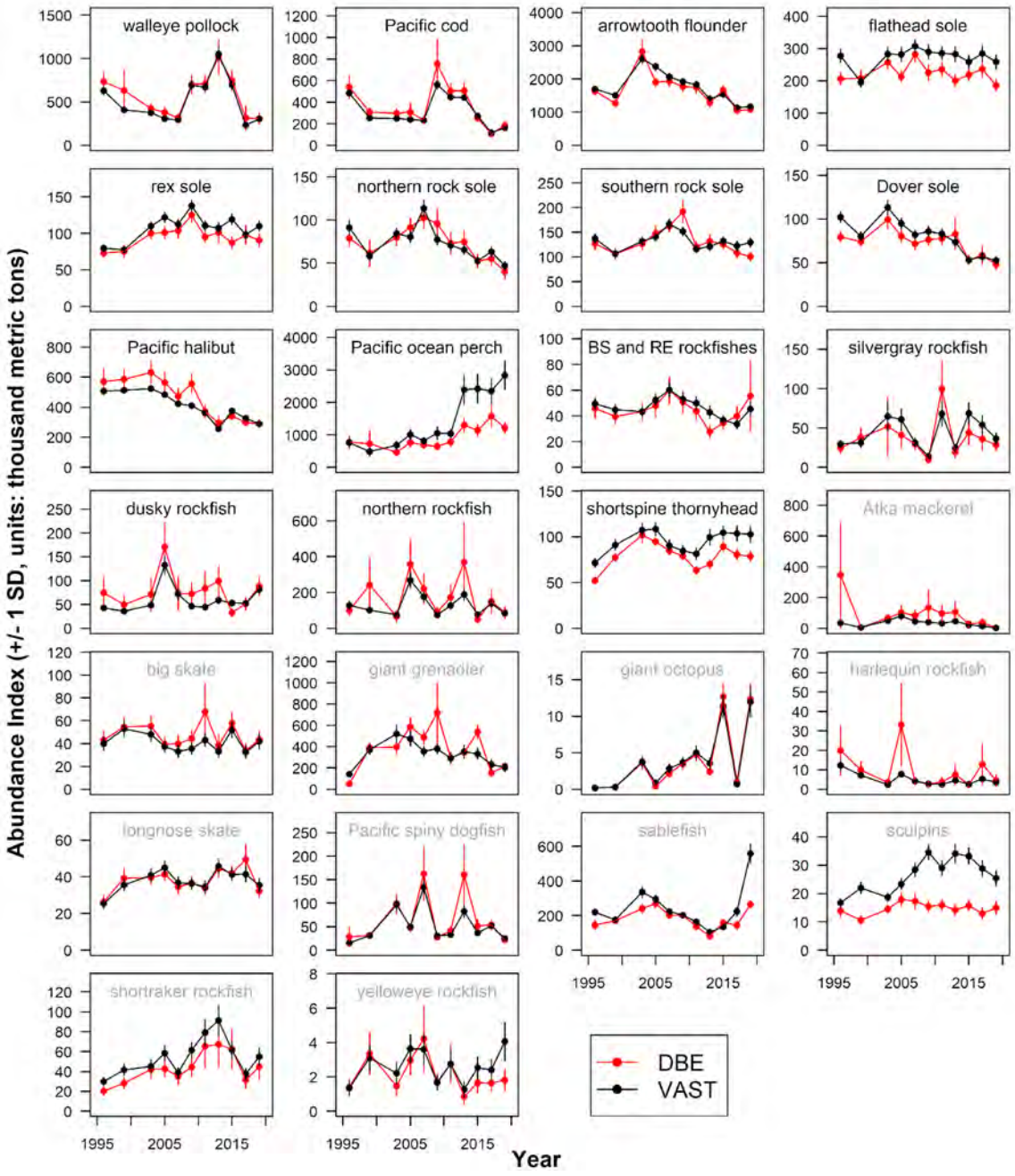


Figure 1. -- Estimates (+/- one standard error) of the index of abundance from 1996 to 2019 for a subset of species (species names in grey were excluded from the survey optimization) using a design-based estimator (DBE; red) and a model-based estimator (vector autoregressive spatiotemporal [VAST] model; black). Model-based estimates shown here are computed with epsilon bias correction for the sake of comparison to design-based estimates.

Single-species Optimizations

Species such as northern and southern rock soles, flathead sole, Pacific cod, walleye pollock, and northern and dusky rockfishes produced single-species survey optimization solutions with higher sampling density in the central and western GOA where they were most abundant, with more stations allocated along the shelf than slope (single-species solutions under scenario A are plotted in Fig. 2). Silvergray rockfish was more restricted to the eastern GOA and thus had an optimal survey design that reflected that distribution, with a high density of samples allocated to the southeastern-most part of the spatial domain (i.e., offshore of Prince of Wales Island). Solutions for deeper-dwelling species like Pacific ocean perch, Dover sole, shortspine thornyhead, and blackspotted/rougheye rockfishes had higher sampling density on the slope.

Stratified random sampling tended to provide lower expected CVs across species than simple random sampling (Fig. 3). There were slight improvements in the expected CV from stratification for many of the flatfishes, regardless of whether the stratification was conducted specifically for individual or multiple species. Expected CVs for species with restricted distributions like silvergray, dusky, and northern rockfishes, Pacific ocean perch, and Atka mackerel were greatly reduced by single-species stratification.

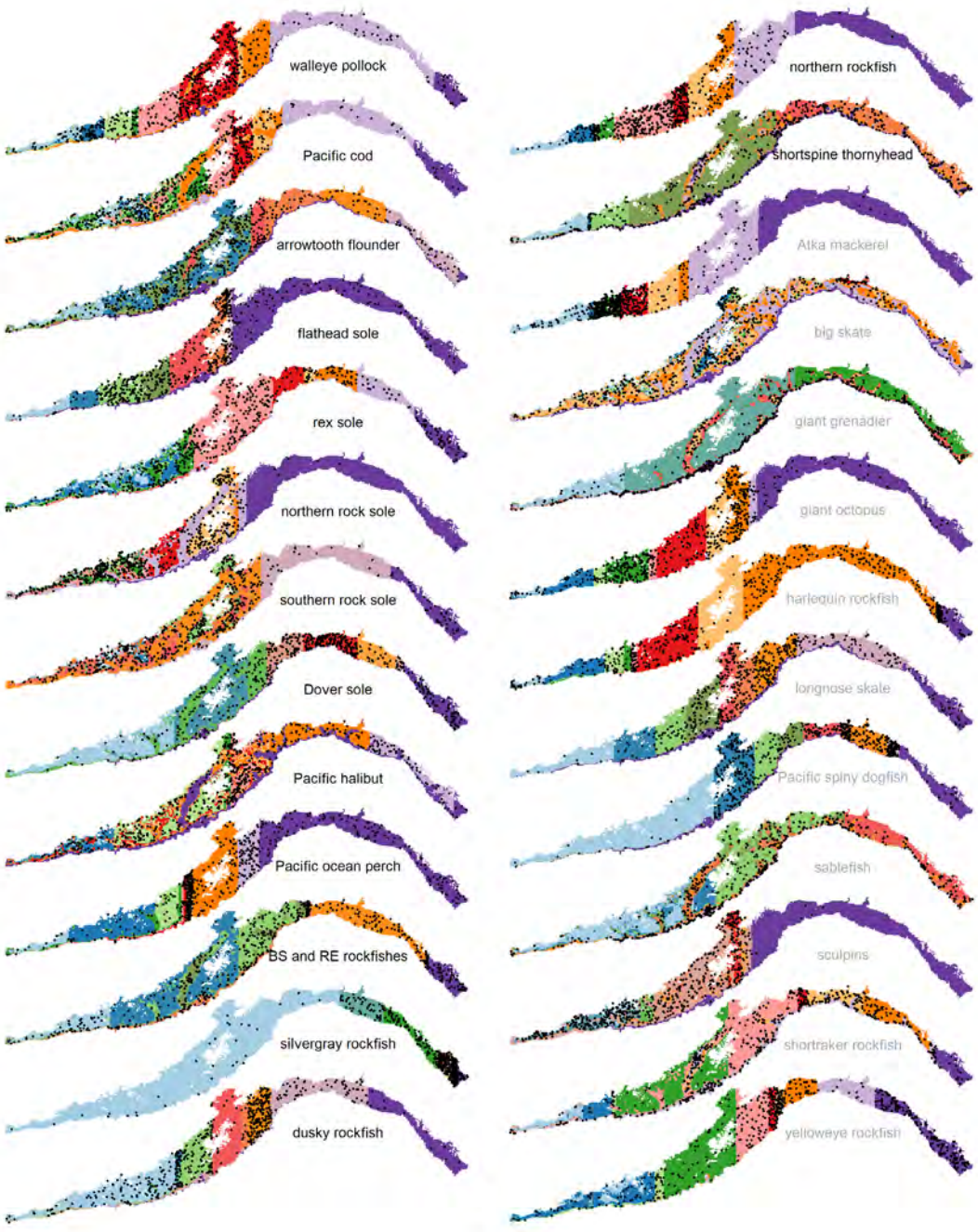


Figure 2. -- Single-species optimized stratified survey solutions conducted under scenario A (10 total strata) plotted on the spatial domain. Different colors indicate different strata. Simulated station locations (black dots) based on the respective solutions (two-boat effort survey, 525 stations) are superimposed to show spatial differences in sampling densities across strata. Species names in grey were non-design species and were excluded from the multispecies survey optimization.

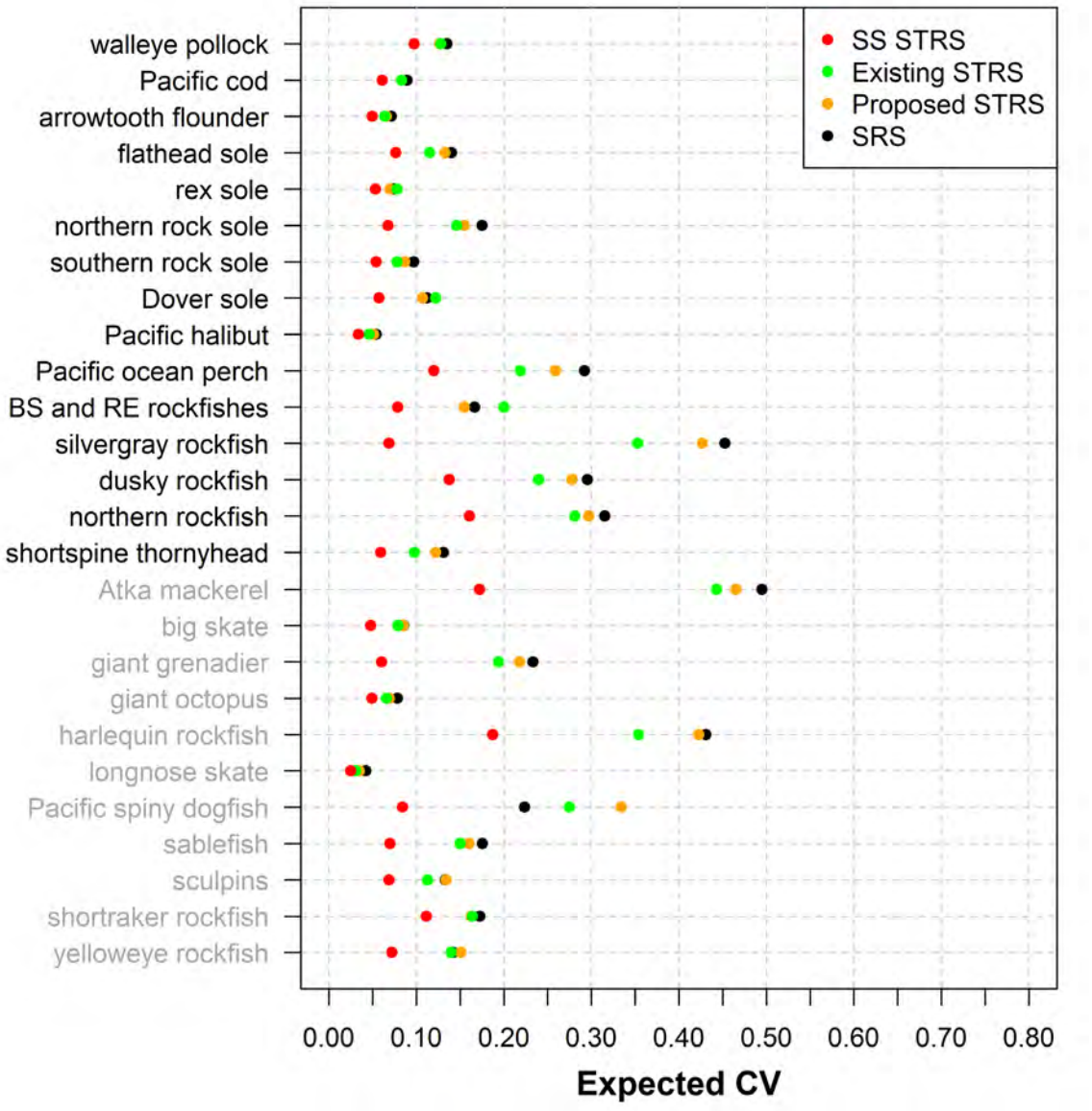


Figure 3. -- Expected coefficients of variation (CVs, Eq. 3) resulting from two-boat effort surveys (525 stations) under various survey designs. Values are shown across species (black: included in the optimization; grey: excluded in the optimization) and survey design type (point color), where SRS: simple random sampling; Proposed STRS: solution corresponding with the area-level (five strata per area, scenario B defined in Table 4) optimization; Existing STRS: existing stratified survey design (scenario L defined in Table 4); SS-STRS: single-species ten-strata gulf-wide optimization (scenario A defined in Table 4).

Multispecies Optimizations

Some features of the single-species stratifications (Fig. 2) were present in the multispecies optimal strata solutions (multispecies solutions under scenarios A and B are shown in Figs. 4 and 5). Shelikof Strait was retained as a stratum in most species' optimal survey designs (Fig. 2). This was also reflected in the presence of Shelikof Strait as a unique stratum in many of the optimized multispecies survey solutions, even when part of Shelikof Strait was divided by the Chirikof and Kodiak management areas in the area-level optimizations (Fig. 4). Consistently across gulf-wide and area-level solutions, there were strata that corresponded to shelf and upper slope areas (e.g., < 300 m) and deeper waters of the slope (Fig. 5). Deep areas were typically assigned to one or two strata for the gulf-wide optimized survey solutions. In contrast, shallower shelf areas were more often defined by additional longitudinal cuts or breakpoints (Fig. 5). Management area boundaries were largely supported as useful strata boundaries, in that gulf-wide solutions demonstrated that longitudinal stratum boundaries were proximal to management area boundaries, even when area boundaries were not explicitly included in the optimization approach (e.g., the boundaries between Yakutat and Southeast, and between Western and Chirikof).

In additional explorations, we explored different settings of the optimization. Of the scenarios defined in Table 4, using depth as the sole stratum variable in a scenario led to the most dramatic changes in the strata boundaries (Appendix B). In the Western management area, scenario B (Appendix B-1) cut the shelf into four longitudinally defined areas, whereas in scenario C (Appendix B-2), the resulting depth stratifications separated a western and eastern shelf area. Both scenarios identified the slope area as one stratum. In the Yakutat, Kodiak, and

Southeast areas, the depth-only stratifications were more concurrent with the existing stratifications, merging many of the existing strata (e.g., slopes and gullies as one stratum, Cook Inlet and Albatross Banks as one stratum, the existing strata in the 1-100 m and 101-200 m depth zones the existing strata as two separate strata in the Southeast region).

Exploratory analyses and discussions with AFSC scientists indicated that a two-boat (550 sample) survey effort scenario was reasonably efficient and the most likely outcome for future surveys. Therefore, we present results from other total effort scenarios in the appendices while comparing options for distributing these 550 samples in the main text. The expected CVs of the proposed optimized (area-level, five strata per area) and existing stratified survey designs generally were between the simple random sample and single-species stratified random survey CVs (Fig. 3). For Dover and rex soles, yelloweye, blackspotted, and rougheye rockfishes, and Pacific spiny dogfish, the expected CVs under the existing survey were higher than those of the simple random design. In contrast, except for Pacific spiny dogfish and yelloweye rockfish, expected CVs of the proposed survey were below those of the simple random design. For the gadids and nearly all flatfish, the expected CVs were below 0.15 regardless of the survey design, indicating that abundance estimates for these species are likely highly informative of trends in biomass in all two-boat (550 station) scenarios. In contrast, the expected CVs of many of the rockfishes were quite high (e.g., CV > 0.25), consistent with the high variability in their distributions and sampled catch rates.

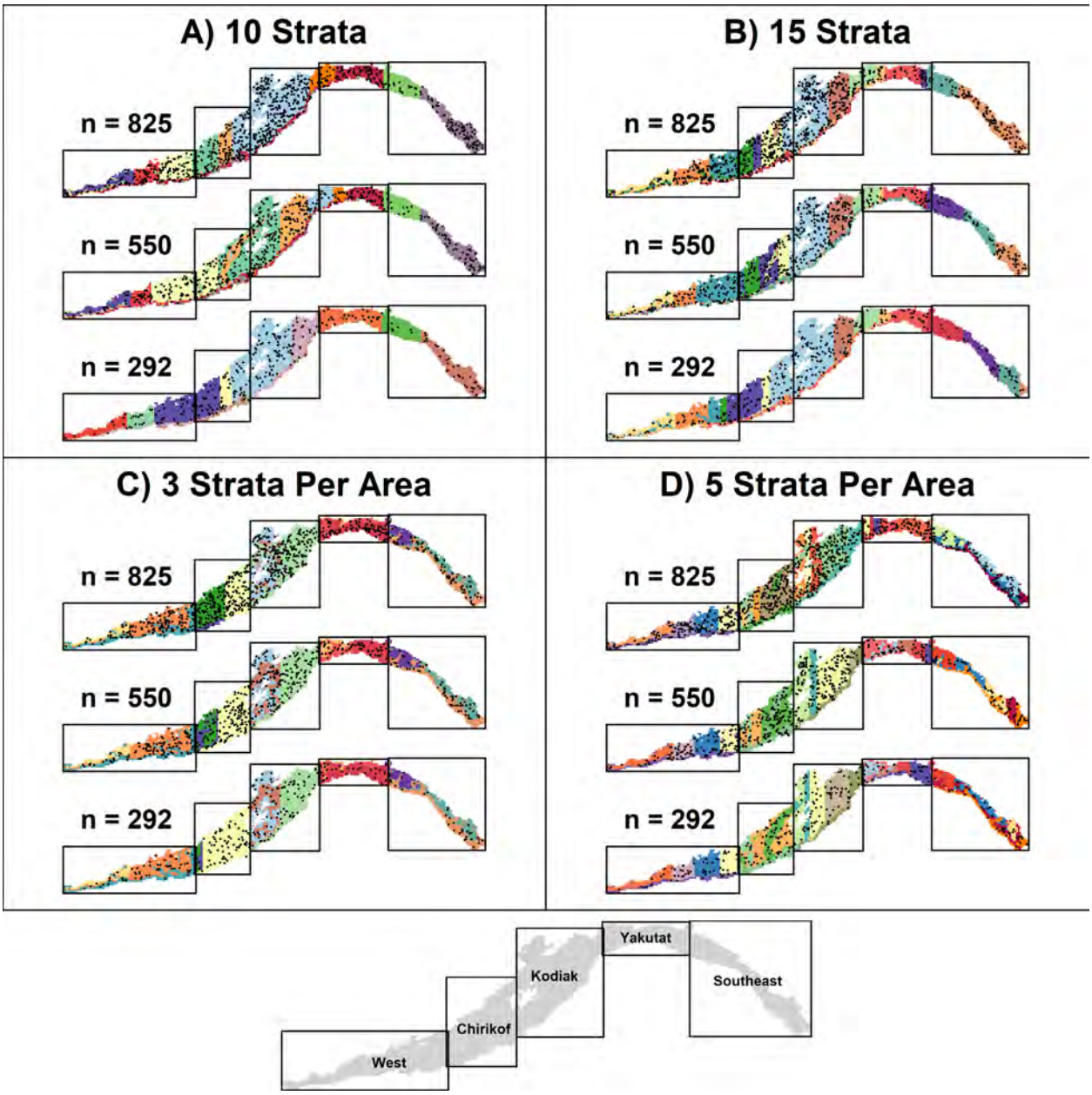


Figure 4. -- Gulf-wide (panels A and B) and area-level (panels C and D) multispecies optimized solutions plotted across the spatial domain. Gulf-wide and area-level optimizations were conducted under scenarios A and B, respectively, as defined in Table 4. Different colors indicate different strata of the proposed designs. The boundaries of the five management areas (labeled at the bottom of the figure) are superimposed.

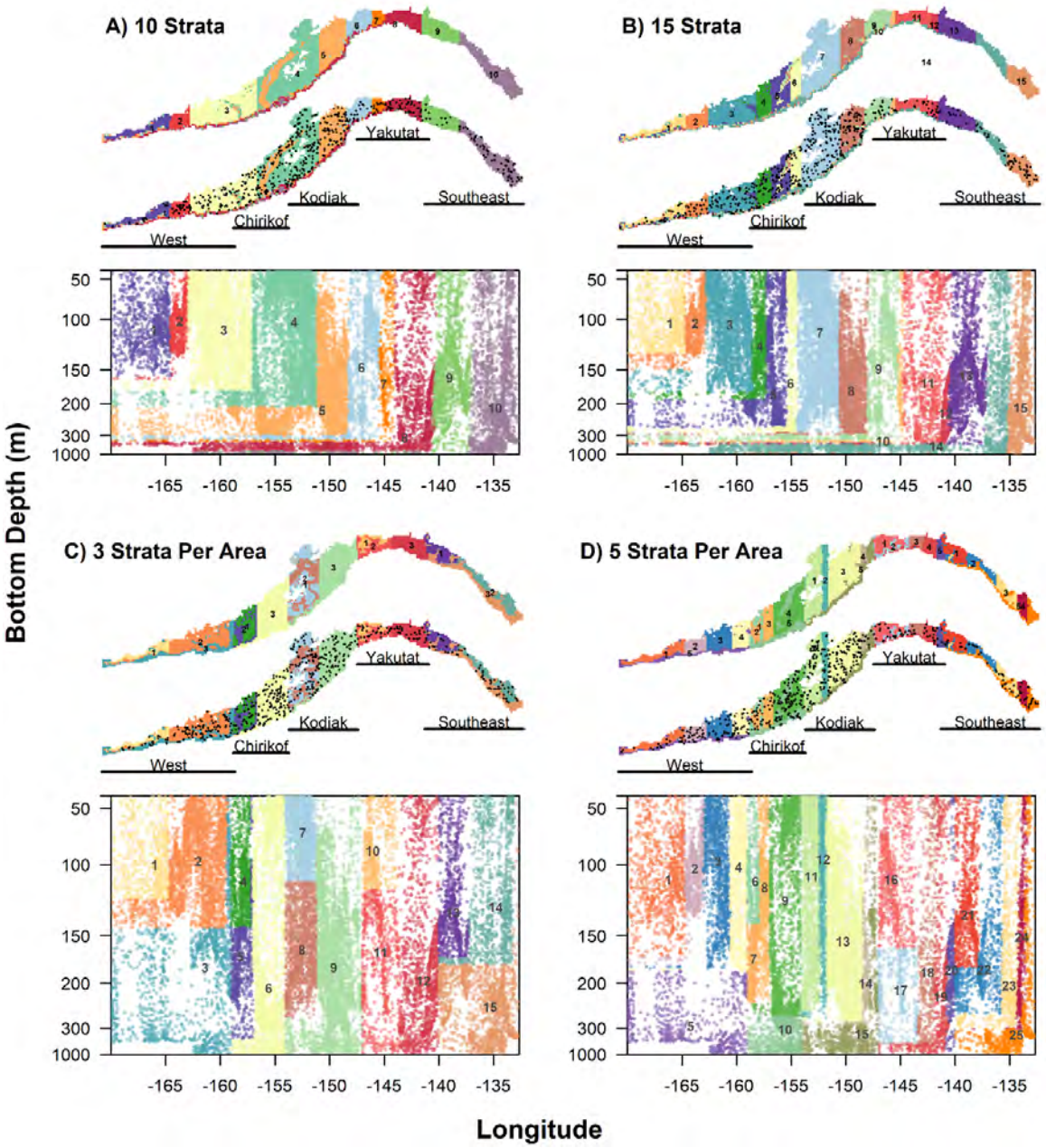


Figure 5. -- Stratum boundary delineations across bottom depth and longitude for the gulf-wide (panels A and B) and area-level (panels C and D) multispecies optimized solutions plotted across the spatial domain. Gulf-wide and area-level optimizations were conducted under scenarios A and B, respectively, as defined in Table 4. The colors designate different strata in the maps (top row of panel) and corresponding stratum variable phase plots (bottom row of panel). The bottom depth axis is scaled to the empirical cumulative distribution function of bottom depths in the prediction domain.

Survey Comparison: Relative Bias

Proposed optimized surveys produced estimates of gulf-wide (Fig. 6) and area-level (Fig. 7) total abundances from simulated surveys with low magnitudes of relative bias. The existing survey design led to abundance estimates with a small bias that was similar to that produced by the gulf-wide optimized design, with the exception of a few species with abundance distributions extending to deeper depths along the slope that showed consistent large negative biases (e.g., Dover sole, shortspine thornyhead, giant grenadier, and sablefish). Relaxing CV constraints for species in areas outside their core distribution (e.g., pollock in the Southeast management area) had little effect on outcomes, and therefore such considerations were not included in developing the area-level solutions. The effort allocation for a two-boat survey design as currently implemented does not sample deep (> 700 m) strata, which contributes to the negative bias of the abundance index for those deeper-ranging species. Trends for the aforementioned species with negative bias for the existing design were consistent for a one-boat design which similarly does not sample deep strata (Appendix C5). However, the negative biases were eliminated when utilizing a three-boat design that includes the deep strata (Appendix C6). Similarly, when the survey was restricted to 700 m and thus the “true” abundances were restricted to those cells shallower than 700 m (scenarios J and K for the proposed designs, and Scenario M for the existing design), the negative biases for those species were also eliminated (Appendix C7).

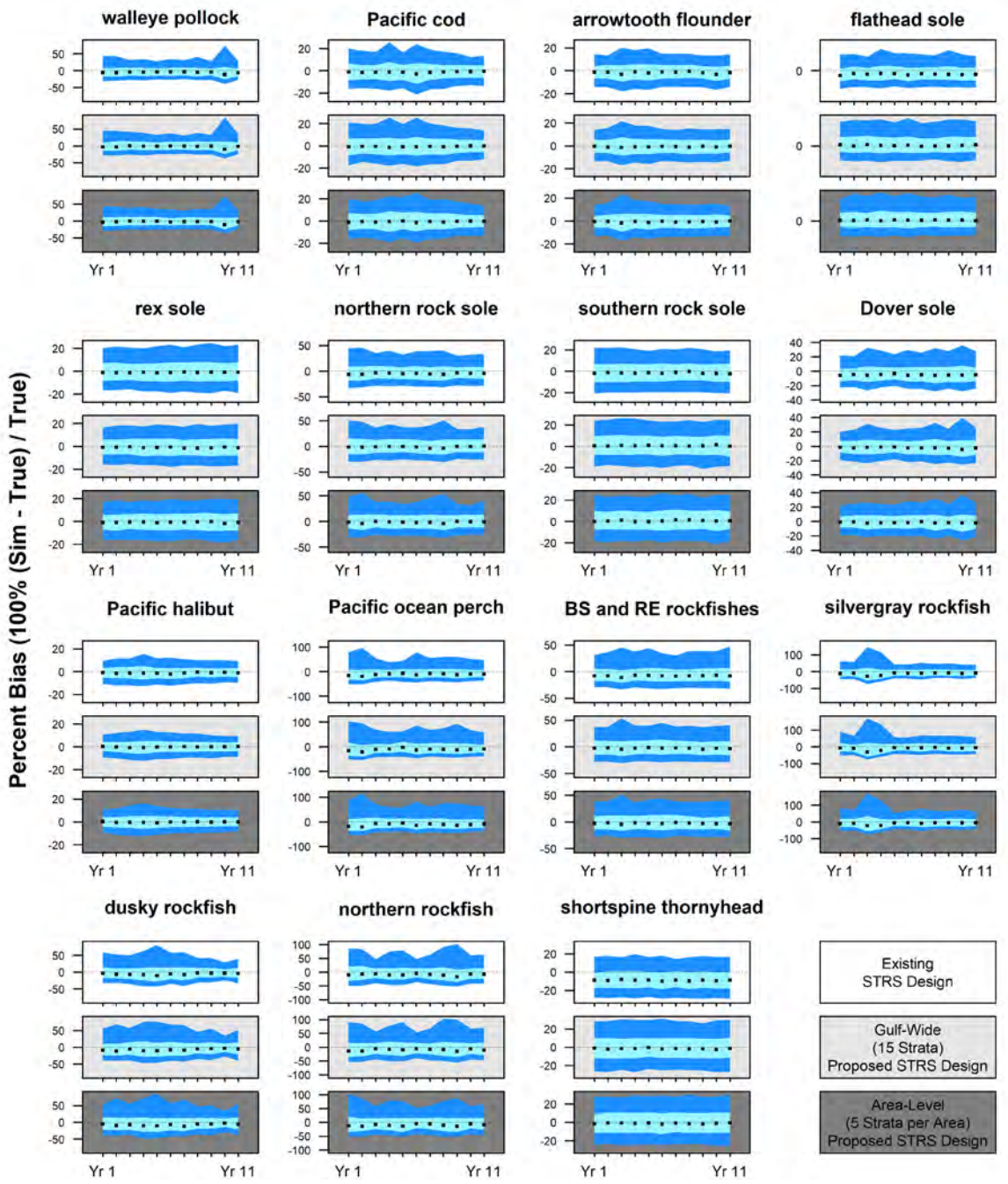


Figure 6. -- Median (with 90% percentile confidence interval) relative bias of simulated total (gulf-wide) abundance indices across time (x-axis) and species. Color rows indicate different surveys: existing survey (white), gulf-wide optimized survey under scenario A defined in Table 4 (light grey, 15 strata), and area-level optimized survey under scenario B defined in Table 4 (dark grey, five strata per area). All survey scenarios were two-boat effort operations for comparison.

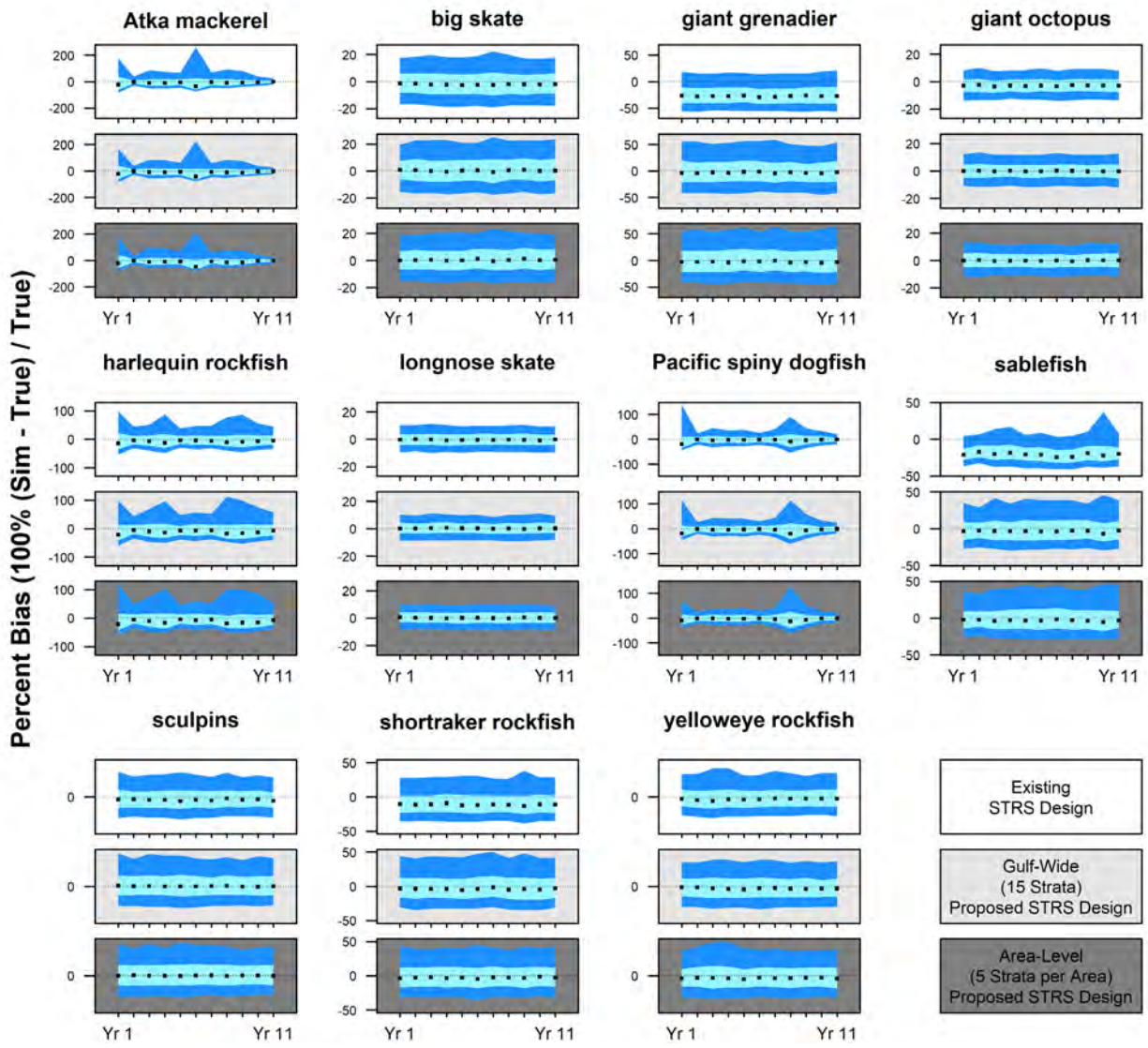


Figure 6. -- Continued.

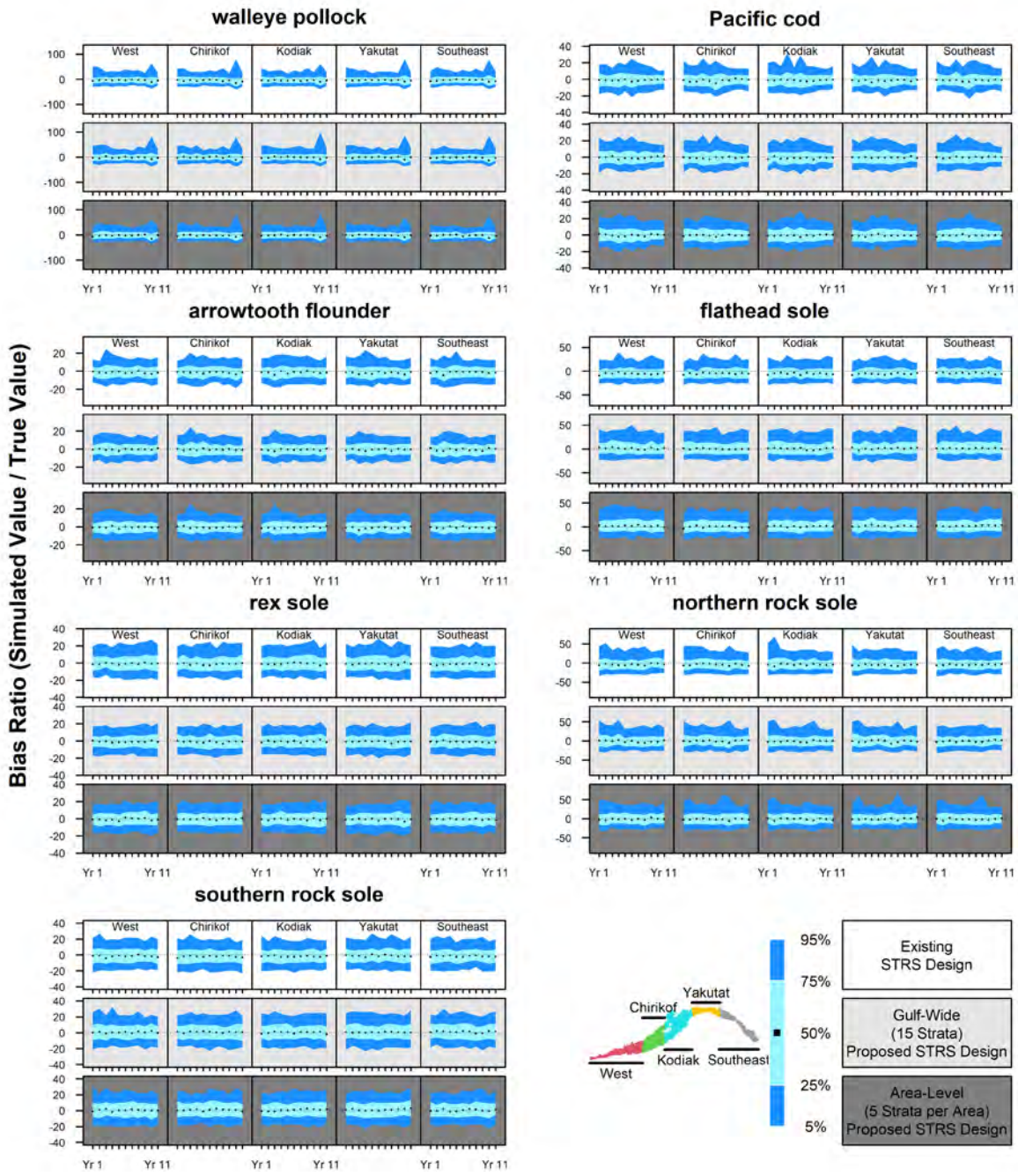


Figure 7. -- Median (with 90% percentile CI) bias ratio of simulated abundance indices across time (x-axis) by management area (column panels) and species. Color rows indicate different surveys: existing survey (white), gulf-wide optimized survey under scenario A defined in Table 4 (light grey, 15 strata), and area-level optimized survey under scenario B defined in Table 4 (dark grey, five strata per area). All survey scenarios were two-boat effort operations for comparison.

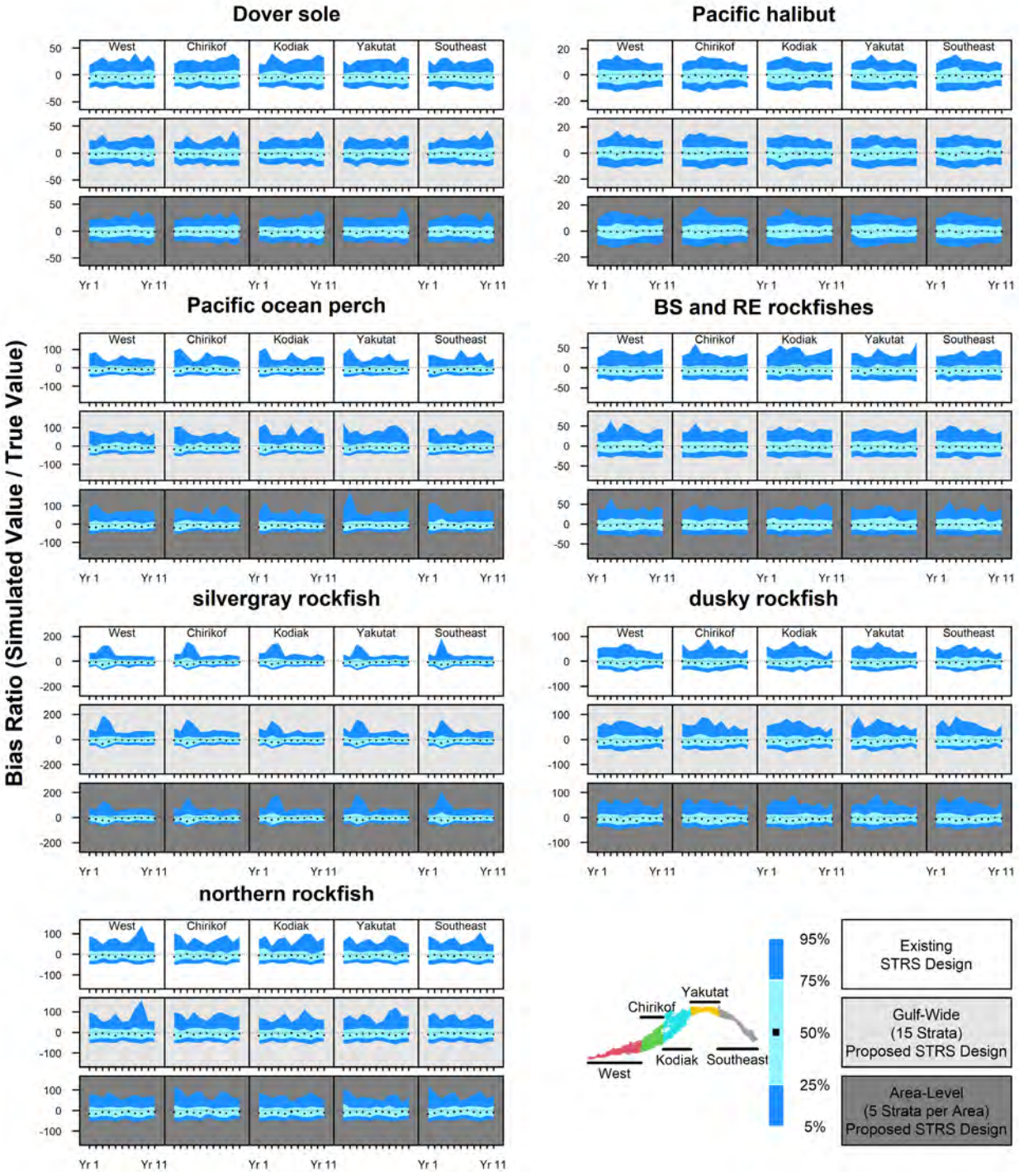


Figure 7. -- Continued.

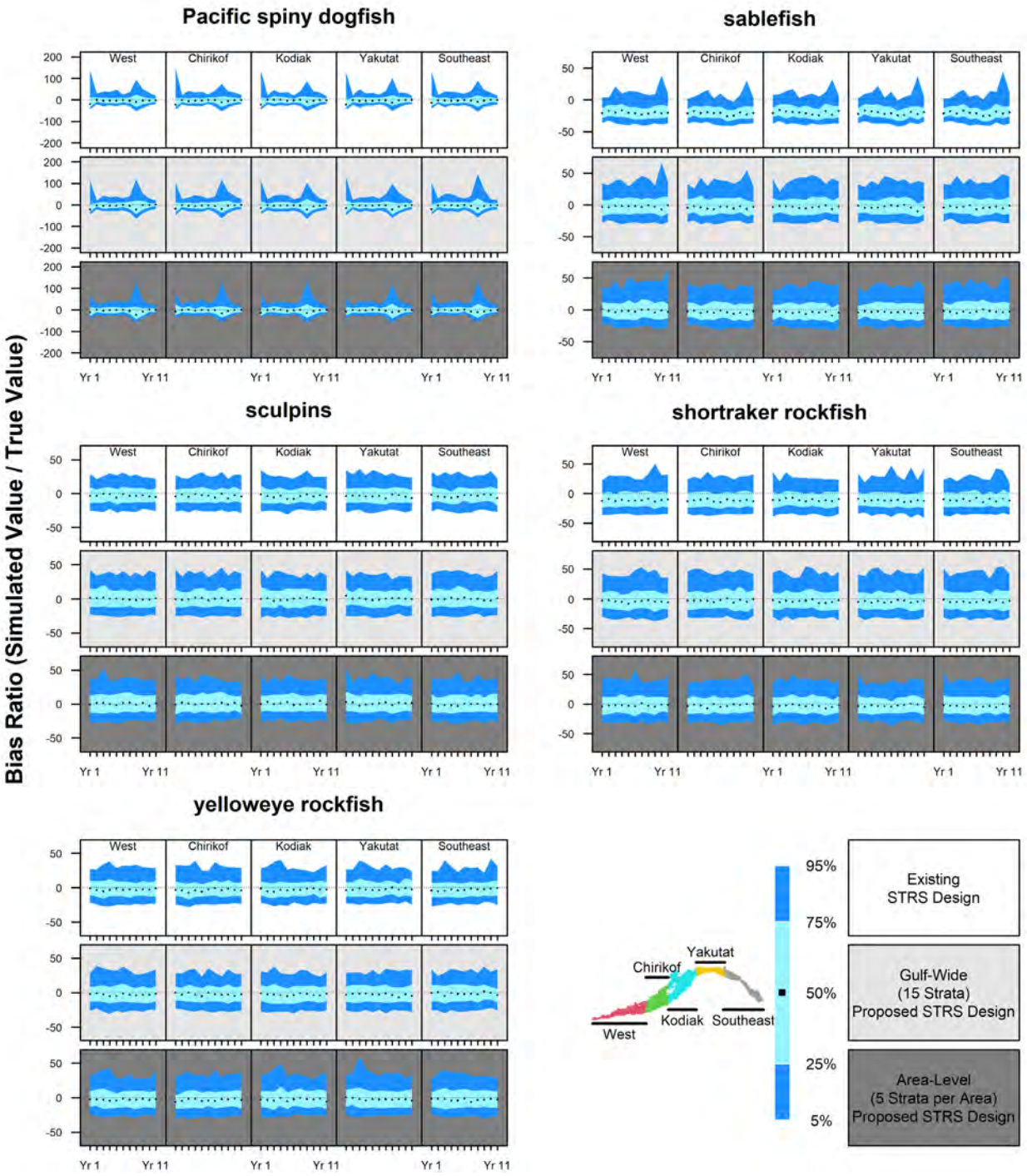


Figure 7. -- Continued.

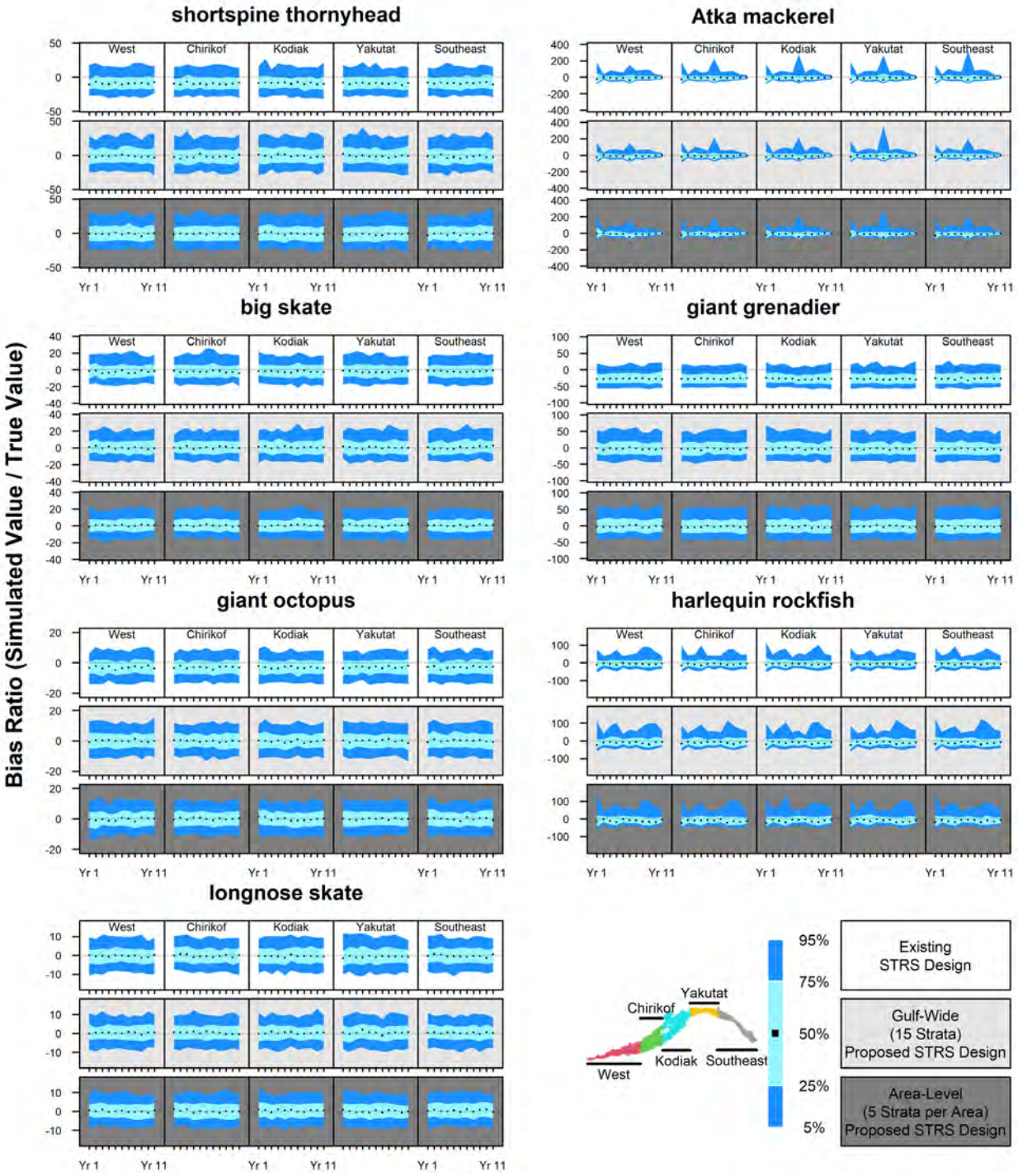


Figure 7. -- Continued.

Survey Comparison: True CV

The existing survey design generally provided true CVs similar to or slightly lower than those of the proposed surveys for most species (Fig. 8). The existing design (scenarios L and M) provided more precise abundance estimates for flathead sole and shortspine thornyhead than all proposed design scenarios (scenarios A-K). However, all proposed design scenarios provided more precise abundance estimates for rex sole, BS and RE rockfishes, and longnose skate. There was little variation in true CV distributions across proposed design scenarios for many of the species included in the optimization. However, there was some variation across proposed design scenarios for species excluded from the survey optimizations like giant grenadier, sculpins, and yelloweye rockfish. The inclusion of depth as the only stratum variable (scenarios C, E, H, I, K) led to higher true CVs and RRMSE of CVs for giant grenadier but lower values for sculpins.

The RRMSE of CV for many of the proposed design scenarios was lower than in the existing design for many species, including shortspine thornyhead, yelloweye, blackspotted and rougheye, and shortraker rockfishes, giant grenadier, big and longnose skates, and Dover and rex soles (Fig. 9). The trends in these results were consistent across scenarios with different levels of simulated sampling error, but the scale of the values differed with greater (e.g., 3-boat, more precise CVs) or lesser (1-boat, less precise CVs) effort (Appendix C1-C4).

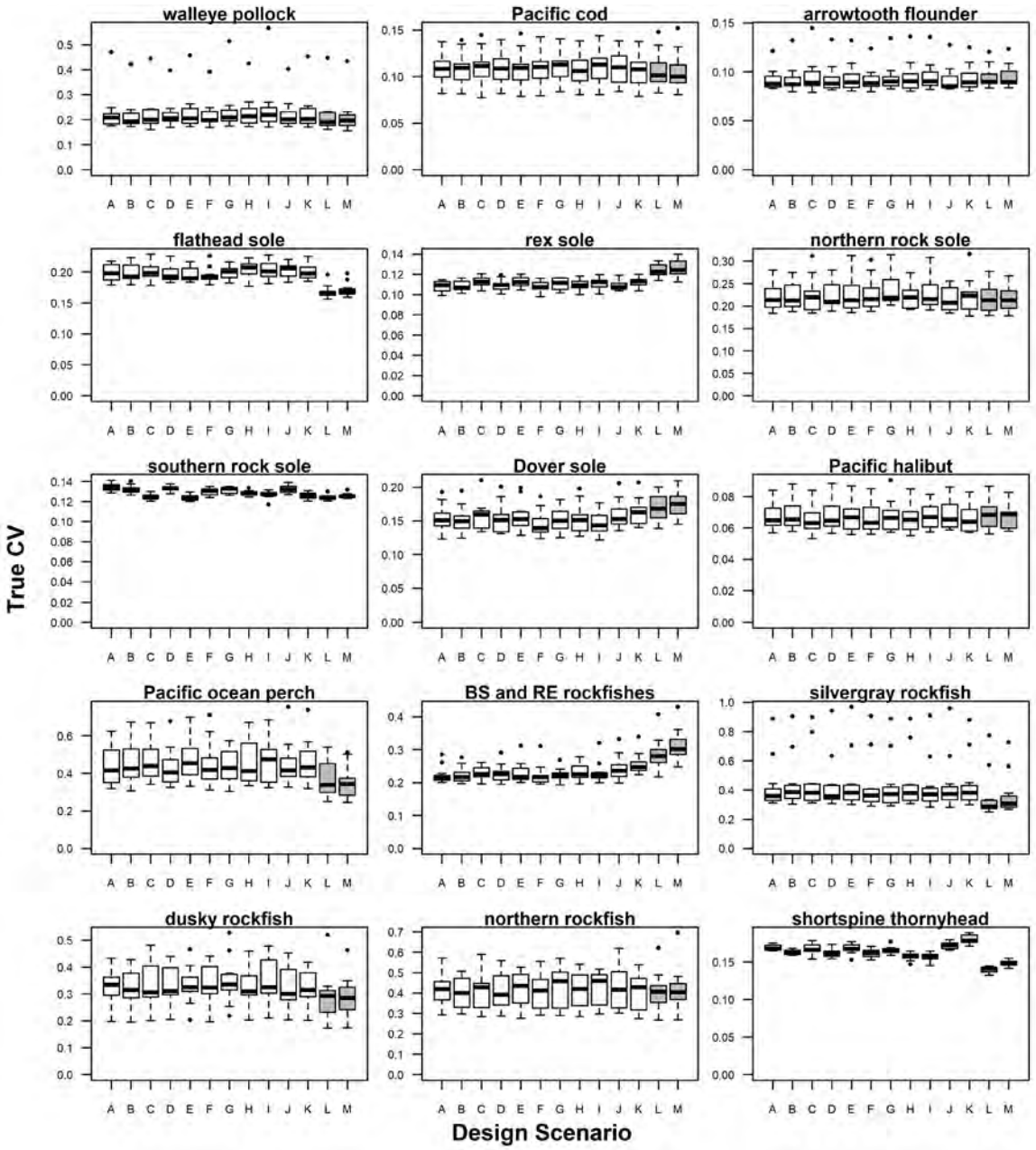


Figure 8. -- Boxplot distributions of true coefficient of variation (CV; standard deviation of estimated abundance indices across 1,000 realized survey simulations divided by the true abundance index) across years for each design scenario described in Table 4. Gulf-scale proposed designs (Scenario A) include 15 strata, area-level designs (Scenarios B-K) include five strata per area, and scenarios L-M utilize the existing survey design. Two-boat operations are shown for comparison. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.

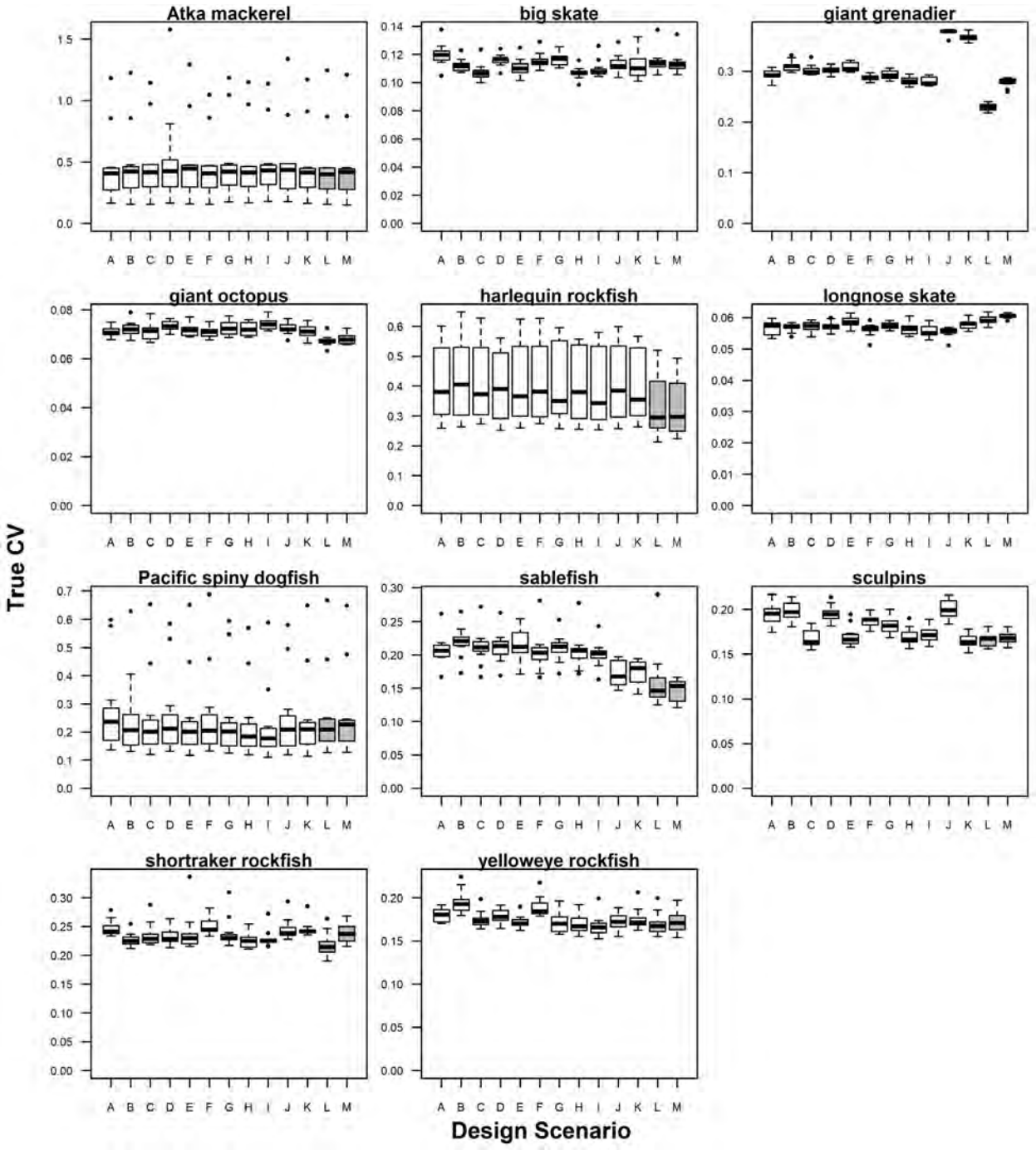


Figure 8. -- Continued.

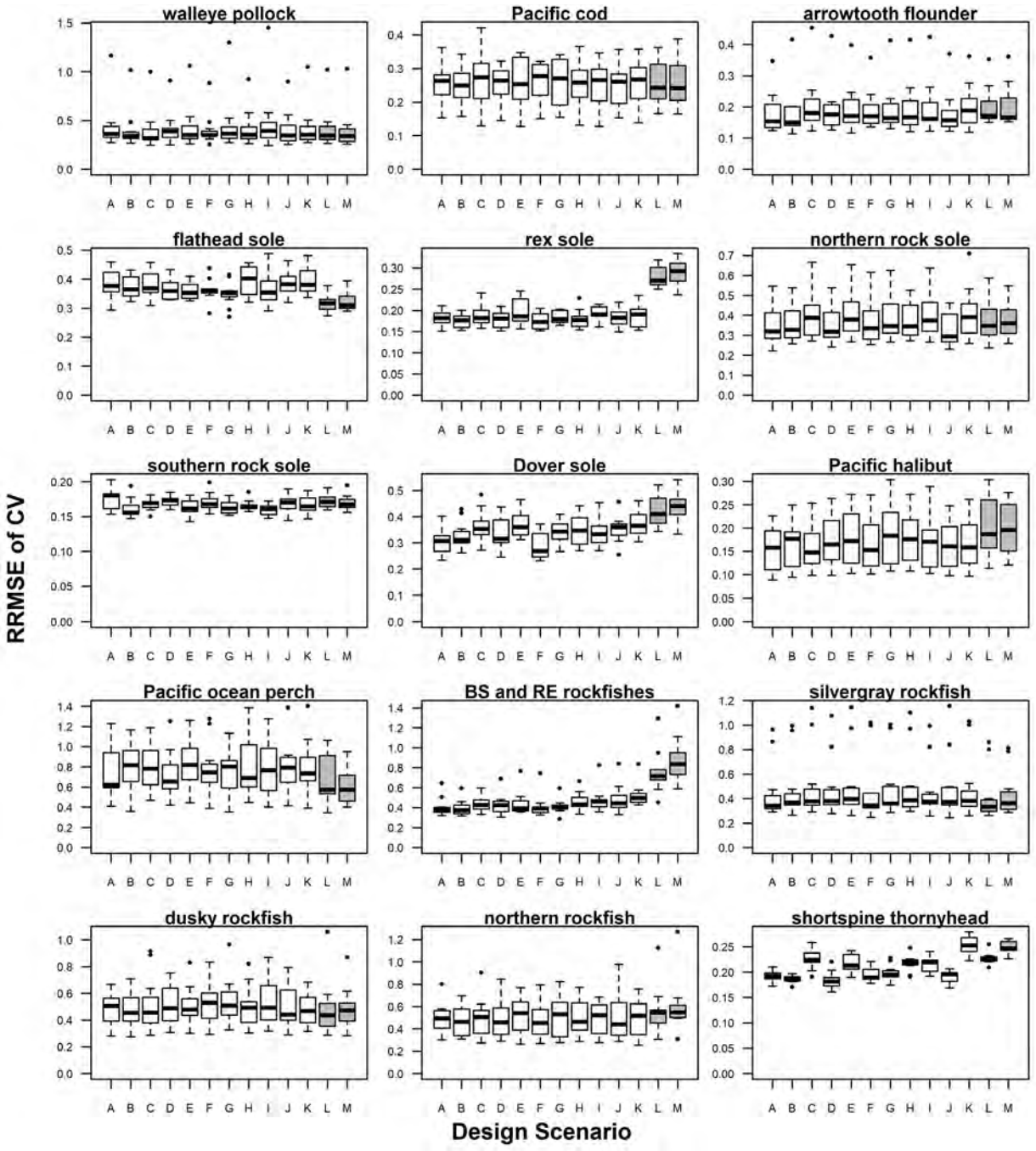


Figure 9. -- Boxplot distributions of the relative root mean square error (RRMSE) of the coefficient of variation (CV) across years for each design scenario described in Table 3. Gulf-scale proposed designs (Scenario A) include 15 strata, area-level designs (Scenarios B-K) include five strata per area, and scenarios L-M utilize the existing survey design. Two-boat operations are shown for comparison. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.

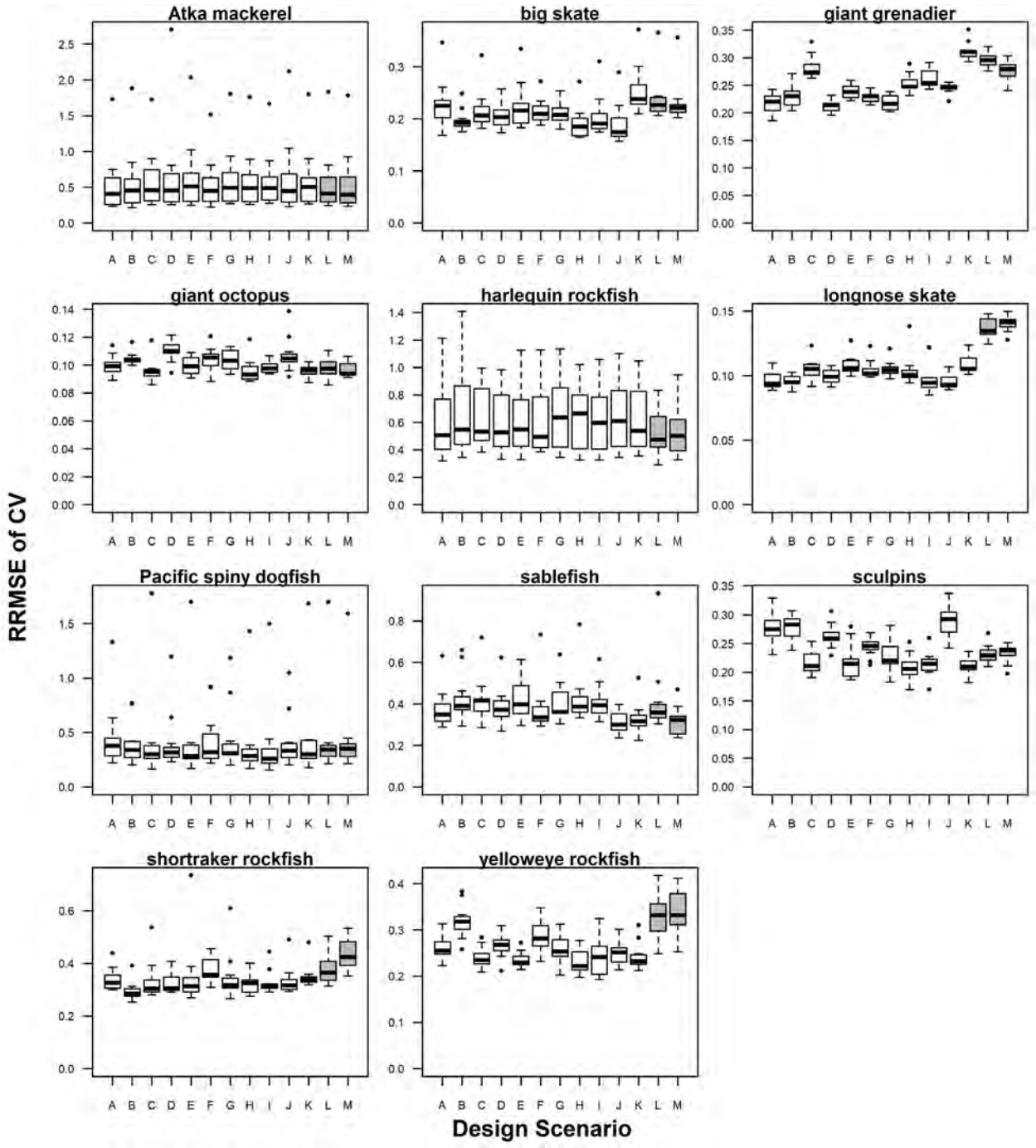


Figure 9. -- Continued.

Lower CV Threshold

Setting an explicit lower CV threshold for all species, below which the optimization no longer seeks to reduce the spatiotemporal variance in abundance, is effective for revealing how to best balance tradeoffs of gains and losses in expected survey performance among species. While the specific value of expected CV required to produce an abundance index with adequate information to inform a stock assessment model is a topic of ongoing research[1], we show the expected response to changing the lower CV threshold from 0 (no threshold) to two potentially reasonable values of 0.15 and 0.25, given area-level optimization solutions with five-strata per area. The results detailed below were similar given alternative thresholds of 0.1 and 0.2. We also evaluated the effects of additional combinations of species-specific CV thresholds, which did not qualitatively change the tradeoffs detailed here, but may warrant further investigation in future efforts.

In the base case with no threshold (or lower CV threshold = 0), rockfishes generally had higher expected CVs than flatfishes and gadids (Fig. 10A). Specifically, expected CVs of Pacific ocean perch and northern, dusky, and silvergray rockfish were approximately 0.25 to 0.4, much higher than all other species, which had expected CVs of approximately 0.05 to 0.20. When the lower CV threshold was set to 0.15 (Fig. 10B), expected CVs decreased for blackspotted and rougheye rockfish and the four aforementioned rockfishes with highly uncertain estimates. The expected CVs slightly increased or remained constant for the species with expected CVs that fell

[1] P.D. Spencer. Alaska Fisheries Science Center, 7600 Sand Point Way NE, Bldg. 4, Seattle, WA 98115. unpubl. data.

below this value under optimizations conducted without a threshold. When the lower CV threshold was further increased to 0.25 (Fig. 10C), the expected CVs for the four rockfishes which previously had expected CVs above this threshold decreased more substantially; however, tradeoffs were more apparent as these increases in precision for species with high uncertainty in abundance estimates led to decreases in precision for flatfishes and gadids, yet the latter set of species were still well below the lower CV threshold.

As the lower CV threshold increased from 0 to 0.25, sample allocation as determined by the Bethel algorithm shifted to adjust to changing objectives among species. Sample allocation increased primarily in the western and Chirikof areas and in deeper strata within the Yakutat and Southeast areas as the allocation algorithm gave greater weight to the rockfishes, some of which are most prevalent in the eastern GOA. To compensate for these increases, allocation largely decreased in the Kodiak and Yakutat areas.

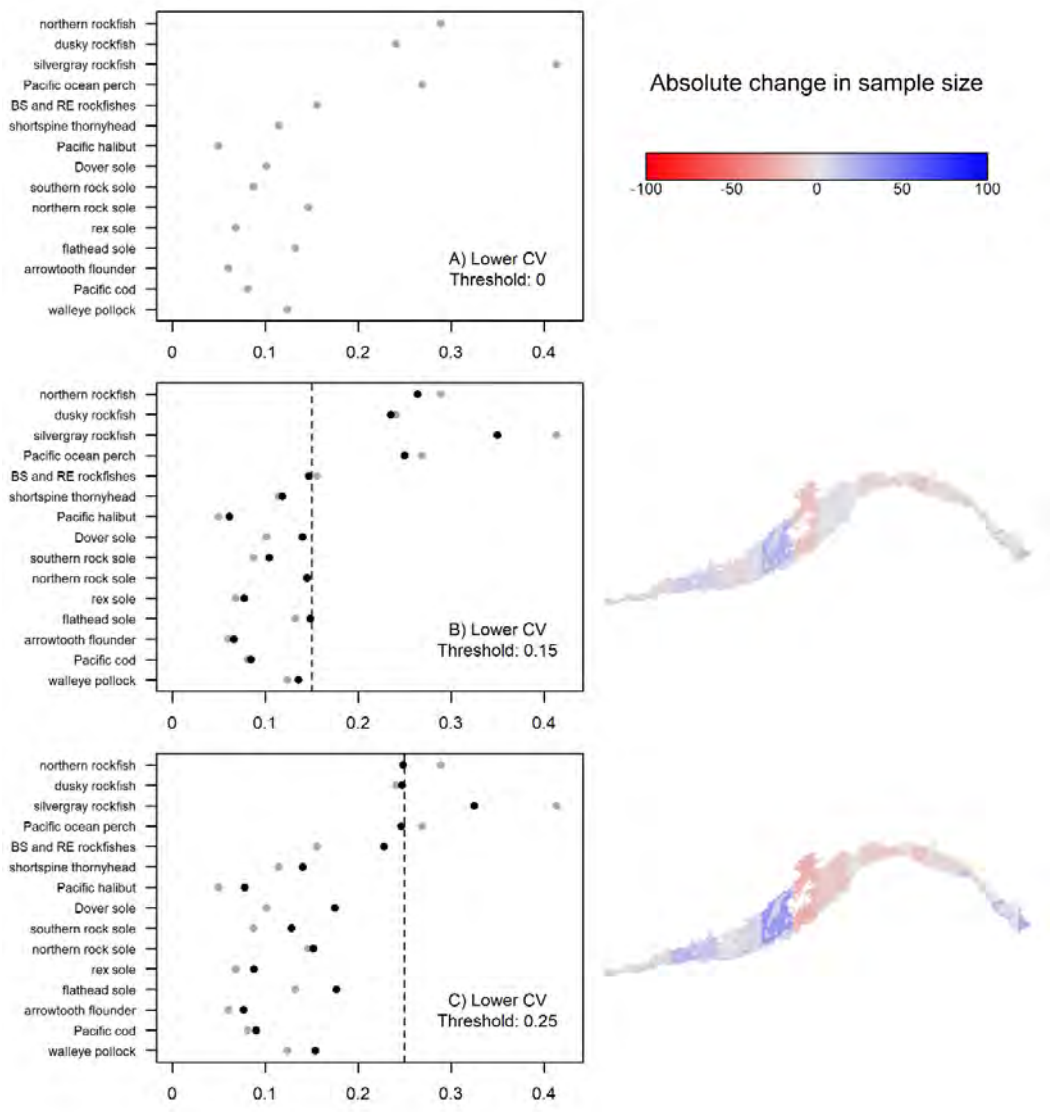


Figure 10. -- Expected coefficients of variation (CV) across species for a two boat-effort (550 stations), area-level (five strata per area, scenario B defined in Table 4) survey optimization under different levels of lower CV thresholds with additional constraints: A) 0 or no explicit threshold, B) 0.15, and C) 0.25. Grey points reproduce the expected CV when the lower CV threshold is zero (thus are constant across panels); black points indicate the expected CV computed for a given CV threshold; vertical lines indicate the CV threshold. If a CV constraint (e.g., $U_1, U_2, \ldots U_G$, Eq. Set 3) for a species is lower than the threshold, then the value of the CV constraint for that species is set to the threshold. The expected CV can still be below the threshold, as the CV constraint only serves as an upper limit. Plotted on the right panels are the corresponding changes in sample allocation across strata relative to the no threshold scenario.

Travel Distances: Existing Versus Proposed Survey Designs

We estimate that the proposed optimized survey design is feasible to implement given the similarity between the cumulative distance traveled by each boat in the existing and proposed surveys. Cumulative distance traveled by each boat in the survey is provided in Table 5. The estimated cumulative distance traveled is not always consistent with the true distance; this is because of the way the algorithm chooses the station order. Compared to historical surveys that used two boats (2011, 2013, 2017, and 2019), the proposed survey design has a slightly larger cumulative distance traveled (8,352 +/- 255 km for boat 1 and 8,340 +/- 257 for boat 2, versus a mean of 7,143 +/- 620 km traveled per boat estimated for historical surveys; Table 5). Estimated distances tend to be shorter than true distances because real-life station choices are made based on proximity, depth, weather, and other factors; the Traveling Salesperson Problem (TSP) solution optimizes based on total distance and thus will likely be shorter than the true distance traveled (See Appendix D for the survey station paths for the historical observations). Cumulative distance for the two-boat optimized survey is slightly longer than the historical cumulative distances because stations are randomly assigned to boats, whereas in historical surveys, stations were sometimes non-randomly assigned. Distances between neighboring stations are also comparable between the historical surveys and the proposed design (Fig. 11).

Table 5. -- Cumulative distance (km) traveled by each boat in the survey, based on the Traveling Salesperson Problem (TSP) solution for station order, with stations assigned randomly to boats. True distance, from historical surveys, is based on the cumulative distance traveled in a survey by each boat (not counting trips to and from port). Approximate shortest cumulative distance is based on the TSP solution. Optimized survey distances are averages from 1,000 realizations of the proposed survey design (area-level, five strata per area, scenario B defined in Table 4).

		Boat 1		Boat 2		Boat 3	
Year	Number of stations	True cumulative distance (km)	Approx. shortest cumulative distance (km)	True cumulative distance (km)	Approx. shortest cumulative distance	True cumulative distance (km)	Approx. shortest cumulative distance
1996	795	10025	7703	10545	8269	8431	7822
1999	757	8198	7019	6734	5778	8245	6741
2003	804	8695	5952	8567	6802	8064	6071
2005	831	12772	6678	10740	6429	8962	5550
2007	805	8971	5304	11631	6917	11248	6267
2009	821	10938	6153	11524	6794	10404	6487
2011	669	12803	7868	11650	6802	--	--
2013	547	11813	7292	10587	6417	--	--
2015	768	12475	7688	11346	6401	9117	5306
2017	534	11564	7309	10210	6278	--	--
2019	541	12129	7995	11872	7183	--	--
Two-boat optimized survey	550	--	8352 (sd: 255)	--	8340 (sd: 257)	--	--

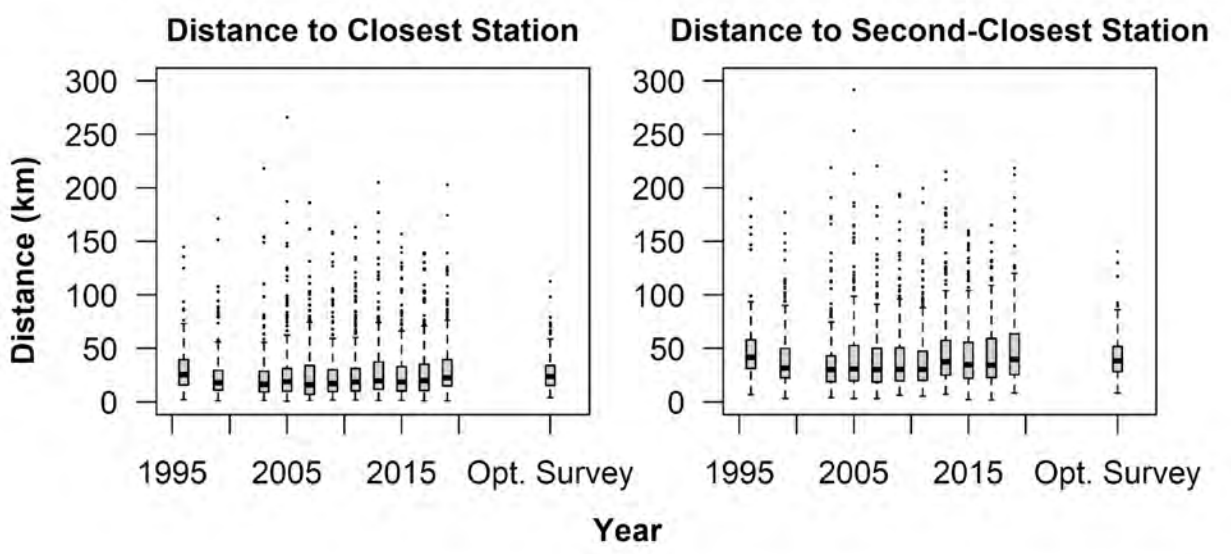


Figure 11. -- Distribution of distances between each station and the two closest unsampled stations, for historical surveys (shown by year at left) and for a single solution from the proposed survey design (area-level, five strata per area, scenario B defined in Table 4). The solution shown here is from an area-level survey optimization with three strata for each area. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.

DISCUSSION

The framework illustrated here is useful for providing strategic and tactical guidance for survey stratification and sample allocation decisions. Additionally, it can be used to inform timely, evidence-based decisions for survey modifications based on quantitative advice, which is a vast improvement over *ad hoc* approaches. The modularity and adaptability of the framework make it flexible and extendable for a range of purposes. For example, one can easily change the assumptions underlying the population density predictions, or update the original predictions in response to observed changes in species distribution and abundance (via the OM). Furthermore, the framework can easily accommodate changes in total sampling effort and the number and spatial extent of strata (via the stratification algorithm). Finally, the spatial allocation of samples among strata can be easily re-optimized to meet a given set of objectives for the precision of abundance estimates (via the sample allocation algorithm). The ability to set lower and upper constraints on the expected precision of abundance indices, whose values can be constant or vary across species, provides a formal framework for addressing prioritization and tradeoffs among species in multispecies surveys.

Practical Recommendations for Survey Design

There were several advantages of the proposed optimized stratified survey designs compared to the existing survey design. Area-level solutions exhibited low levels of bias in both gulf-wide and area-level abundance indices and thus may be useful for stock assessment and management purposes (e.g., total allowable catch apportionment). These solutions were also

shown to be as logistically feasible as the existing survey design in terms of the total distance traveled between stations given the same sample size. Although true CVs were lower in the current survey design relative to the proposed design for many of the deeper-ranging species, their corresponding abundance estimates were often negatively biased (due to missing population densities in deeper strata) and had higher uncertainty and bias in the CV estimates given the existing survey design.

While we illustrate many potential survey optimization solutions here to provide a range of options for survey planning, it is necessary to identify the most desirable of the proposed designs upon which to conduct further investigation or modification. For example, since solutions were similar across magnitudes of sampling effort (Fig. 4), the two-boat effort (550 stations) survey design can be utilized and adjusted for different levels of effort using the Bethel algorithm to adjust the sampling allocation to the total number of survey stations that can be sampled given the time and resources available in a given year. This reduced version of the optimization can be done with relative ease and lower computation time while remaining consistent with the optimization framework. This was the strategy employed when considering different CV thresholds to better balance the optimization gains by tailoring allocation to reduce CVs of the species with greater uncertainty in abundance estimates (Fig. 10). The level of the lower CV threshold could be species-specific, something for which this optimization framework can easily accommodate. Ongoing research will inform the choice of the CV threshold values based on how survey index precision affects bias and uncertainty of biomass estimates from stock assessment models (Spencer et al. in prep, ICES 2020). A cost-benefit analysis evaluating the relationship between total sampling effort, precision, and downstream management quantities

such as total allowable catch can more directly address multispecies tradeoffs of surveys as they relate to management advice and ultimately the economic value of fisheries (Francis 2006).

The lower CV threshold analysis highlights another difference between our proposed survey design and the existing survey design in how species are prioritized and whether such prioritization can be adjusted to suit management needs. The flexibility to prioritize improvements in biomass index precision of specific species under the existing design is unclear, but likely limited to possibly modifying the species weights used in the Neyman sample allocations. The existing survey design prioritizes more economically important and abundant species in the effort allocation scheme (von Szalay and Raring 2018), whereas the lower CV threshold analysis focuses on achieving more precise estimates for species with greater uncertainty, which results in placing more weight on rockfish species in the optimal solutions by relaxing CV constraints for species whose abundance is already precisely estimated. Further interaction with the North Pacific Fishery Management Council and Plan Teams will need to address whether this is considered a desirable outcome from a management perspective. Regardless, the methods proposed here provide a framework to facilitate such adjustments to species prioritization based on management needs, whereas it is unclear how such adjustments can or should be performed in the existing design and allocation approach. In our analysis, the mechanism to reduce CVs across species while understanding the limitations and tradeoffs is straightforward because it can be evaluated by changing the user-defined CV constraints as an input to the optimization. Thus, our approach allows for both long-term strategic planning based on stock assessment model needs and short-term tactical adjustments to changes in survey resources or species of particular management concern in a given year.

Contrasting expected outcomes from simple random sampling (SRS) and stratified random surveys optimized for each species to those optimized across species is insightful for determining both the influence of each species on multispecies optimization solutions and the limit to improvements in estimation that can be gained from sampling design alone. Single-species optimizations represent a scenario in which all resources are utilized to obtain an optimal stratified survey for an individual species in the absence of interspecific trade-offs. Together with the expected SRS CV, the range of feasible expected precisions can be calculated, which is useful in communicating the interspecific tradeoffs of a multispecies survey optimization (i.e., in the context of the minimum expected CV for a given effort level). For example, a majority of the species exhibited relatively precise expected CVs (e.g., < 0.15) regardless of survey design type, with stratification providing only slight increases in precision. This was the case for many broadly distributed and economically important species. In contrast, rockfishes (e.g., silvergray, blackspotted and rougheye, northern, and dusky rockfishes) exhibited large increases in precision with single-species survey optimization. These large precision improvements are partly due to the patchier and restricted spatial distribution of many rockfishes (e.g., the western-dominant distribution for northern rockfish and eastern-dominant distribution for silvergray rockfish). Thus, strata and allocation of stations are optimized to reflect those distributions in the single-species solutions. These gains in precision relative to the expected SRS CVs for these species are not realized to that extent in the multispecies optimizations (Fig. 4) due to the tradeoffs that occur when the optimization incorporates many species with diverse spatial distributions, which also change over time. For these rockfishes, precise estimates (i.e., CV between 0.15 and 0.20) can only be achieved under a single-species optimization framework. Furthermore, these

analyses can help inform difficult decisions regarding which species should be deprioritized if improvements in precision are inadequate to produce informative inputs to stock assessments. This clarifies the limitations and consequences of maintaining a multispecies, multi-objective survey.

Caveats: Effects of Untrawlable Habitat and Varying Catchability

The measure we used to express the true CV encompasses the expected sampling variability of a population estimate across multiple realized survey simulations. While we demonstrated how differences in spatial availability (attributable to whether the design sampled the complete distribution of abundance by depth, particularly in the deepest extent of the OM) and observation error contributed to true CV, there are likely other sources of variation in sampling efficiency present in the historical and simulated data that are not explicitly considered here. These include measurement error and spatial variation in gear efficiency due to density-dependent population processes, unaccounted environmental processes, and habitat effects that could not be quantified due to the presence of untrawlable habitat. For example, we assumed that all cells in the spatial domain were available for sampling when in reality, a considerable portion of the GOA is untrawlable by the BTS gear due to the complex terrain, high-rugosity rocky reefs, steep slopes, and strong currents (Baker et al. 2019). The data used to inform the OM came from locations in the spatial domain that were sampled by the survey gear. Thus, it was assumed that density in untrawlable areas could be interpolated from a model with observations from only trawlable areas. An additional related assumption applies to the design-based estimator in that fish densities are assumed not to differ between untrawlable and trawlable cells. This assumption

is likely not appropriate for many species, given interspecific differences in species-habitat associations. For example, untrawlable habitats may be more likely to have higher densities of rockfishes and lower densities of flatfishes given their typical hard substrates with high rugosity (Jones et al. 2012, 2021). In summary, the non-random spatial distribution of untrawlable areas unaccounted for in our OM survey simulations and the assumption of constant catchability in space and time (Kotwicki and Ono 2019, Cordue 2007) are likely significant sources of both additional bias and sampling variability.

Future Considerations Regarding Timescales of Ecosystem Change and Survey Adaptation

It is important to carefully consider three interconnected aspects of time scale that may influence future survey performance: 1) how quickly species abundances and distributions are changing; 2) how far back in time the historical data input to the OM should extend; and 3) whether and how frequently survey designs should be adjusted given new observations. Using all the years of observations available integrates the broadest range of frequencies and magnitudes of temporal, spatial, and spatiotemporal variation. However, given the fitted spatiotemporal distributions of the past may not be reflective of the future, especially under novel environmental conditions (Muhling et al. 2020), it is worth considering using only survey data from more recent years to condition the OM and inform short-term decisions (Ault et al. 1999). In this case, it may also be pertinent to frequently update the design given optimizations incorporating survey observations collected after the initial optimization. While this can cause problems for the interpretation of long-term time series given some designs, according to sampling theory,

stratified random designs produce inherently unbiased estimates and therefore are more flexible to changes in stratification and sample allocation. Ratios of estimated abundance calculated under proposed and existing designs near unity (Appendix C-7) further support the consistency of the time series to changes in STRS designs, with the caveat that the survey footprint is consistent between designs. Given the existing two-boat survey design, this meant restricting the survey spatial domain to those cells less than 700 m.

We envision that one frontier of adaptive monitoring of fishery resources may include updating survey design or sample allocation based on near-term predictions of future species distributions. If species distributions can be skillfully forecasted in the near term using dynamic environmental covariates, stratified surveys can be updated to account for expected future species distribution shifts. For example, in the eastern Bering Sea, the spatial extent of the cold pool (e.g., 2ºC isotherm) has been shown to be predictive of groundfish abundance and distribution (Thorson 2019b, Kotwicki and Lauth 2013, Spencer 2008). Given the temporal variability of the cold pool extent (e.g., Thorson 2019b), survey designs could be re-optimized to adjust to forecasted changes in the distributions of the species due to the expected extent of the cold pool. Further evaluation would be needed to determine how such an adaptive approach would influence survey efficiency relative to a static approach similar to that described here.

CONCLUSIONS

We presented a formalized stratified random survey optimization of the GOA multispecies BTS that constrains spatiotemporal variance in abundance. The optimization allows

for species to be prioritized based on precision, allowing more flexibility to modify the survey design to match species-specific prioritizations based on management and stock assessment needs. Proposed solutions from the optimization produced expected precisions of abundance estimates comparable to the existing STRS design but with higher accuracy and lower bias. Simulated surveys under the proposed designs also had similar total distances travelled between stations than the existing survey design, indicating that the proposed design is likely feasible to implement with the same number of sampling days per survey season.

This optimization approach addresses many practical challenges for fishery-independent multispecies surveys in the Alaska region and beyond by providing a consistent and objective framework for quantifying interspecific tradeoffs and mitigating consequences of reductions in sampling effort due to fluctuations in survey resources. Survey effort reduction often requires changes to sampling design, and we propose the use of the framework outlined here to determine how to adjust when faced with such challenges. The consequences of historical and proposed survey effort reductions on the quality of biomass estimates in this and other regions in Alaska, and the best approach to mitigating them, will also be informed by ongoing efforts related to variation in sampling intensity and survey footprint (ICES 2020, von Szalay et al. in prep), and by comparison of effects on both design- and model-based estimators[2]. The techniques developed here will likely be applied to existing Alaska bottom trawl surveys with stratified random designs in other regions (e.g., eastern Bering Sea slope survey) and to potentially reallocate samples from other survey designs where needed to increase spatial coverage.

[2] E.A. Laman, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Bldg. 4, Seattle, WA 98115. unpubl. data.

ACKNOWLEDGMENTS

We thank all staff of the Alaska Fisheries Science Center's (AFSC) Groundfish Assessment Program (GAP), program affiliates, volunteers, and charter captains and crew for collecting the data analyzed here during the GOA Groundfish Bottom Trawl Survey. We also thank the GOA and Aleutian Islands Team within GAP and the NPFMC GOA Groundfish Plan Team for their constructive feedback on earlier versions of this analysis. We are grateful to Cole Monnahan, Jim Ianelli, Michael Martin, and Ned Laman for comments that improved the report. Finally, we thank Jodi Pirtle, Ned Laman, and Mark Zimmermann for their contributions to creating and providing the best available bathymetry data to fit the scope of this project. This research was performed while ZSO held an NRC Research Associateship award at the Alaska Fisheries Science Center.

CITATIONS

Ault, J. S., Diaz, G. A., Smith, S. G., Luo, J., and J. E. Serafy. 1999. An efficient sampling survey design to estimate pink shrimp population abundance in Biscayne Bay, Florida. N. Am. J. Fish. Manage. 19:696-712. doi:https://doi.org/10.1577/1548-8675(1999)019<0696:AESSDT>2.0.CO;2.

Baker, M. R., Palsson, W., Zimmermann, M., and C. N. Rooper. 2019. Model of trawlable area using benthic terrain and oceanographic variables - Informing survey design and habitat maps in the Gulf of Alaska. Fish. Oceanogr. 28(6):629-657. doi:https://doi.org/10.1111/fog.12442.

Banerjee, S., Gelfand, A. E., Finley, A. O., and H. Sang. 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70: 825-848. doi:https://doi.org/10.1111/j.1467-9868.2008.00663.x.

Barcaroli, G. 2014. SamplingStrata: An R Package for the Optimization of Stratified Sampling. J. Stat. Softw. 61(4):1–24. doi: https://www.jstatsoft.org/article/view/v061i04.

Cao, J., Chen, Y., Chang, J. H., and X. Chen. 2014. An evaluation of an inshore bottom trawl survey design for American lobster (*Homarus americanus*) using computer simulations. J. Northwest Atl. Fish. Sci. 46(27):39. doi: https://doi.org/10.2960/J.v46.m696.

Cochran, W.G. 1977. Sampling techniques, 3rd edition. John Wiley & Sons, New York, NY. 428 p.

Cordue, P. L. 2007. A note on non-random error structure in trawl survey abundance indices. ICES J. Mar. Sci. 64:1333–1337. doi: https://doi.org/10.1093/icesjms/fsm134.

Cressie, N., and C. K. Wikle. 2011. Statistics for spatio-temporal data. John Wiley & Sons, New York, NY. 624 p.

Fissel, B., Dalton, M., Felthoven, R., Garber-Yonts, B., Haynie, A., Himes-Cornell, A., Kasperski, S., Lee, J., Lew, D., and C. Seung. 2019. Stock assessment and fishery evaluation report for the groundfish fisheries of the Gulf of Alaska and Bering Sea/Aleutian Islands area: Economic status of the groundfish fisheries off Alaska, 2018. North Pacific Fishery Management Council, 1007 West Third, Suite 400, Anchorage, Alaska 99501.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68(6):1124-1138. doi: https://doi.org/10.1139/f2011-025.

Hartig, F. 2020. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.3.3.0. https://CRAN.R-project.org/package=DHARMaICES, 2020.

Hahsler, M., and K. Hornik, 2020. TSP: Traveling Salesperson Problem (TSP). R package version 1.1-10. https://CRAN.R-project.org/package=TSP.

Hilborn, R., and C. J. Walters. 2013. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media, New York, NY. 570 p. doi:http://doi.org/10.1007/978-1-4615-3598-0.

ICES Workshop on unavoidable survey effort reduction (WKUSER). ICES Scientific Reports. 2:72, 92 pp. doi:http://doi.org/10.17895/ices.pub.7453.

Johnson, K. F., Thorson, J. T., and A. E. Punt. 2019. Investigating the value of including depth during spatiotemporal index standardization. Fish. Res. 216:126-137. doi:https://doi.org/10.1016/j.fishres.2019.04.004.

Jones, D., Wilson, C. D., De Robertis, A., Rooper, C., Weber, T., and J. L. Butler. 2012. Evaluation of rockfish abundance in untrawlable habitat: combining acoustic and complementary sampling tools. Fish. Bull., U.S. 100(3):332–343.

Jones, D. T., Rooper, C. N., Wilson, C. D., Spencer, P. D., Hanselman, D. H., and R. E. Wilborn. 2021. Estimates of availability and catchability for select rockfish species based on acoustic-optic surveys in the Gulf of Alaska. Fish. Res. 236:105848. doi:https://doi.org/10.1016/j.fishres.2020.105848.

Kotwicki, S., and R. R. Lauth. 2013. Detecting temporal trends and environmentally-driven changes in the spatial distribution of bottom fishes and crabs on the eastern Bering Sea shelf. Deep-Sea Res. Part II 94():231-243. doi:https://doi.org/10.1016/J.DSR2.2013.03.017.

Kotwicki, S., and K. Ono. 2019. The effect of random and density-dependent variation in sampling efficiency on variance of abundance estimates from fishery surveys. Fish Fish. 20:760-774. doi: https://doi.org/10.1111/faf.12375.

Laman, E. A., Harris, J. P., Pirtle, J. L., Conrath, C. L., Hurst, T. P., and C. N. Rooper. In prep. Advancing model-based Essential Fish Habitat descriptions for North Pacific species. Available from Alaska Fisheries Science Center, 7600 Sand Point Way NE, Bldg. 4, Seattle, WA 98115.

Lindgren, F., Rue, H., and J. Lindström. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. B 73:423-498. doi: https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Mueter, F. J., and B. L. Norcross. 2002. Spatial and temporal patterns in the demersal fish community on the shelf and upper slope regions of the Gulf of Alaska. Fish. Bull., U.S. 100(3):559-581.

Muhling, B. A., Brodie, S., Smith, J. A., Tommasi, D., Gaitan, C. F., Hazen, E. L., Jacox, M. G., Auth, T. D., and R. D. Brodeur. 2020. Predictability of species distributions deteriorates under novel environmental conditions in the California Current System. Front. Mar. Sci. 7. doi:https://doi.org/10.3389/fmars.2020.00589.

North Pacific Fishery Management Council (NPFMC). 2020. Fishery management plan for groundfish of the Gulf of Alaska. North Pacific Fishery Management Council, 1007 West Third, Suite 400, Anchorage, Alaska 99501.

Overholtz, W., Jech, J., Michaels, W., Jacobson, L., and P. Sullivan. 2006. Empirical comparisons of survey designs in acoustic surveys of Gulf of Maine-Georges Bank Atlantic herring. J. Northwest Atl. Fish. Sci. 36:127-144. doi: https://doi.org/10.2960/J.v36.m575

Oyafuso, Z. S., Barnett, L. A. K., and S. Kotwicki. 2021. Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty, ICES J. Mar. Sci. fsab038. doi: https://doi.org/10.1093/icesjms/fsab038.

Rooney, S. C., Rooper, C. N., Laman, E. A., Turner, K. A., Cooper, D. W., and M. Zimmermann. 2018. Model-based essential fish habitat definitions for Gulf of Alaska groundfish species.
U. S. Dep. Commer. NOAA Tech. Memo. NMFS-AFSC-373, 370 p.

Rosenkrantz, D. J., R. E. Stearns, and P. M. Lewis. 1977. An analysis of several heuristics for the traveling salesman problem. SIAM J. Comput. 6(3):563–581. doi:https://doi.org/10.1137/0206041.

Rue, H., Martino, S., and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J. R. Stat. Soc. B 71:319-392. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Spencer, P. D. 2008. Density-independent and density-dependent factors affecting temporal changes in spatial distributions of eastern Bering Sea flatfish. Fish. Oceanogr. 17(5):396-410. doi: https://doi.org/10.1111/j.1365-2419.2008.00486.x.

Thorson, J. T. 2017. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. Can. J. Fish. Aquat. Sci. 75(9): 1369-1382. doi: https://doi.org/10.1139/cjfas-2017-0266.

Thorson, J. T., and L. A. K. Barnett. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. 74:1311–1321. doi:https://doi.org/10.1093/icesjms/fsw193.

Thorson, J. T. 2019a. Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210:143-161. doi: https://doi.org/10.1016/j.fishres.2018.10.013.

Thorson, J. T. 2019b. Measuring the impact of oceanographic indices on species distribution shifts: The spatially varying effect of cold-pool extent in the eastern Bering Sea. Limnol. Oceanogr. 64(6):2632-2645. doi: https://doi.org/10.1002/lno.11238.

Thorson, J. T., Cunningham, C. J., Jorgensen, E., Havron, A., Hulson, P. J. F., Monnahan, C. C., and P. von Szalay. 2021. The surprising sensitivity of index scale to delta-model assumptions: Recommendations for model-based index standardization. Fish. Res. 233:105745. doi: https://doi.org/10.1016/j.fishres.2020.105745.

von Szalay, P. G., Kotwicki, S., Barnett, L. A. K., and L. J. Rugolo. In prep. A comparison of alternatives to the traditional stratified random sampling design in the Gulf of Alaska bottom trawl survey: design vs. model-based approaches. Available from Alaska Fisheries Science Center, 7600 Sand Point Way NE, Bldg. 4, Seattle, WA 98115.

von Szalay, P. G., Wilkins, M. E., and M. M. Martin. 2008. Data report: 2007 Gulf of Alaska bottom trawl survey. U. S. Dep. Commer. NOAA Tech. Memo. NMFS-AFSC-189, 247 p.

von Szalay, P. G., and D. A. Somerton. 2017. A method for predicting trawlability in the Gulf of Alaska with the use of calibrated, split-beam, echosounder backscatter. Fish. Bull., U. S. 115(4):496-504. doi:https://doi.org/10.7755/FB.115.4.6.

von Szalay, P. G., and N. W. Raring. 2018. Data report: 2017 Gulf of Alaska bottom trawl survey. U. S. Dep. Commer. NOAA Tech. Memo. NMFS-AFSC-374, 260 p.

Zimmermann, M. 2006. Benthic fish and invertebrate assemblages within the National Marine Fisheries Service US west coast triennial bottom trawl survey. Cont. Shelf Res. 26(8):1005-1027. doi: https://doi.org/10.1016/j.csr.2006.02.013.

Zimmermann, M., and M. M. Prescott. 2014. Smooth sheet bathymetry of Cook Inlet, Alaska. U. S. Dep. Commer. NOAA Tech. Memo. NMFS-AFSC-275, 32 p.

Zimmermann, M., and M. M. Prescott. 2015. Smooth sheet bathymetry of the central Gulf of Alaska. U. S. Dep. Commer. NOAA Tech. Memo. NMFS-AFSC-287, 54 p.

Zimmermann, M., Prescott, M. M., and P. J. Haeussler. 2019. Bathymetry and geomorphology of Shelikof Strait and the Western Gulf of Alaska. Geosciences 9(10):409. doi:https://doi.org/10.3390/geosciences9100409.

APPENDICES

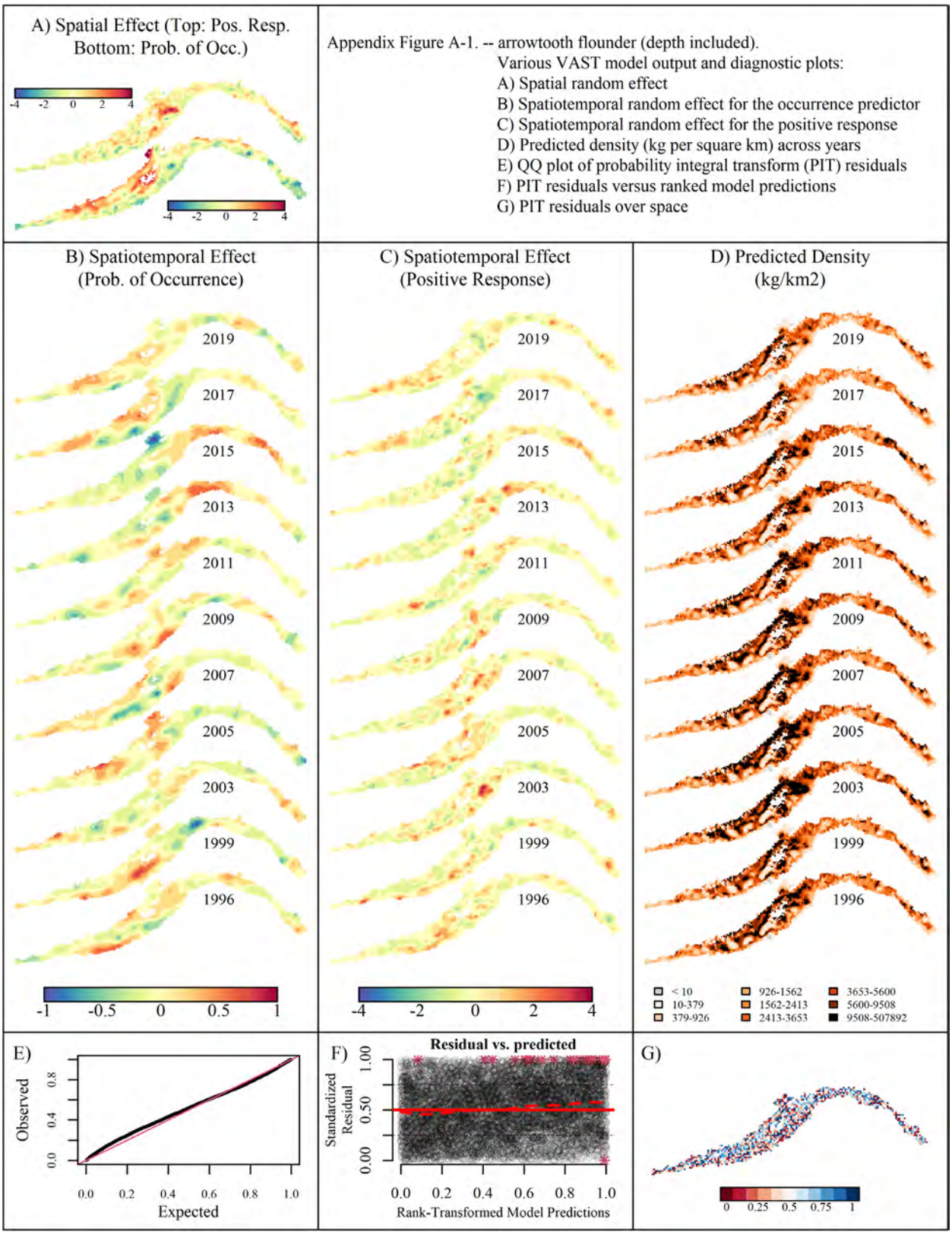
APPENDIX A

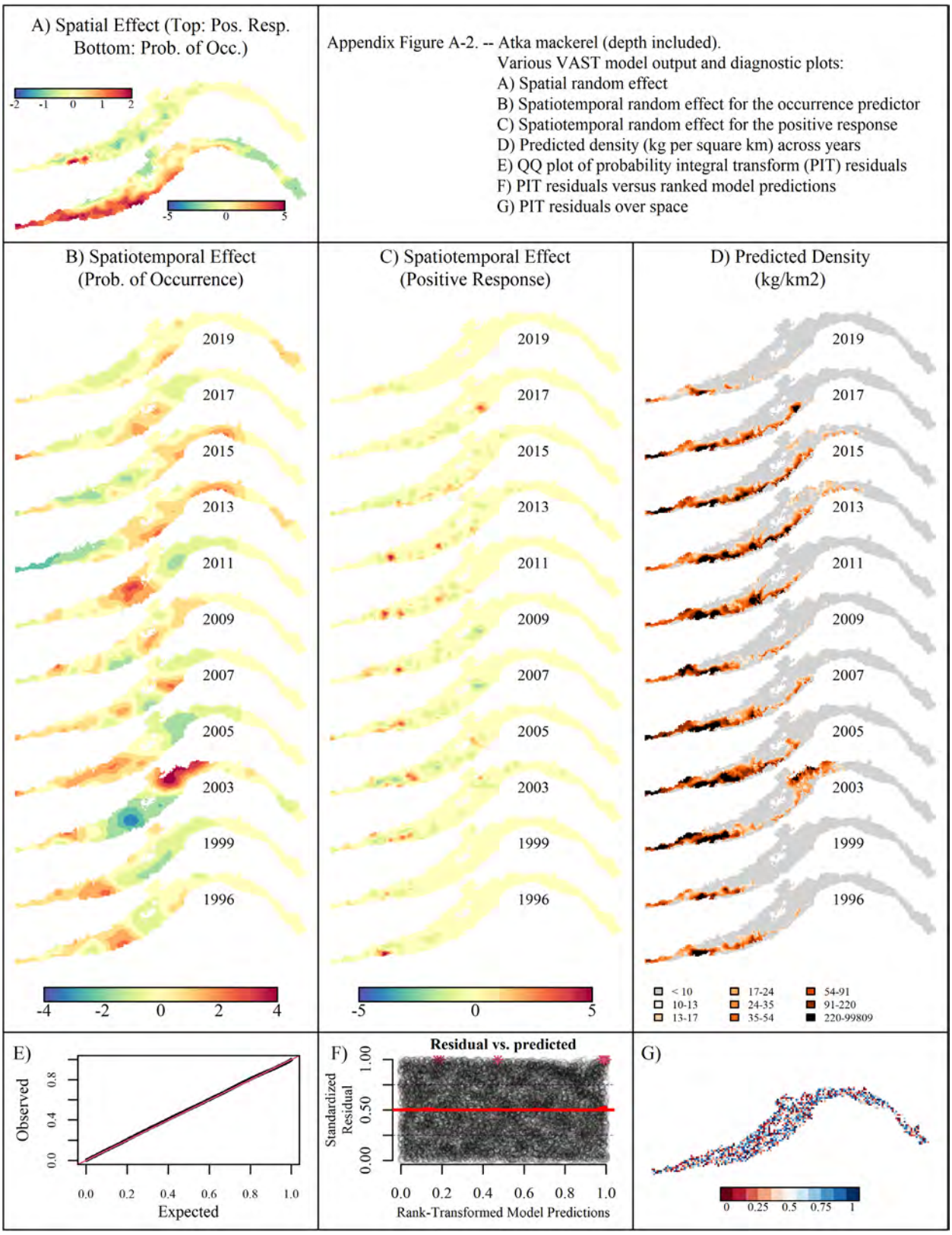
Operating Model Outputs and Diagnostics Plots

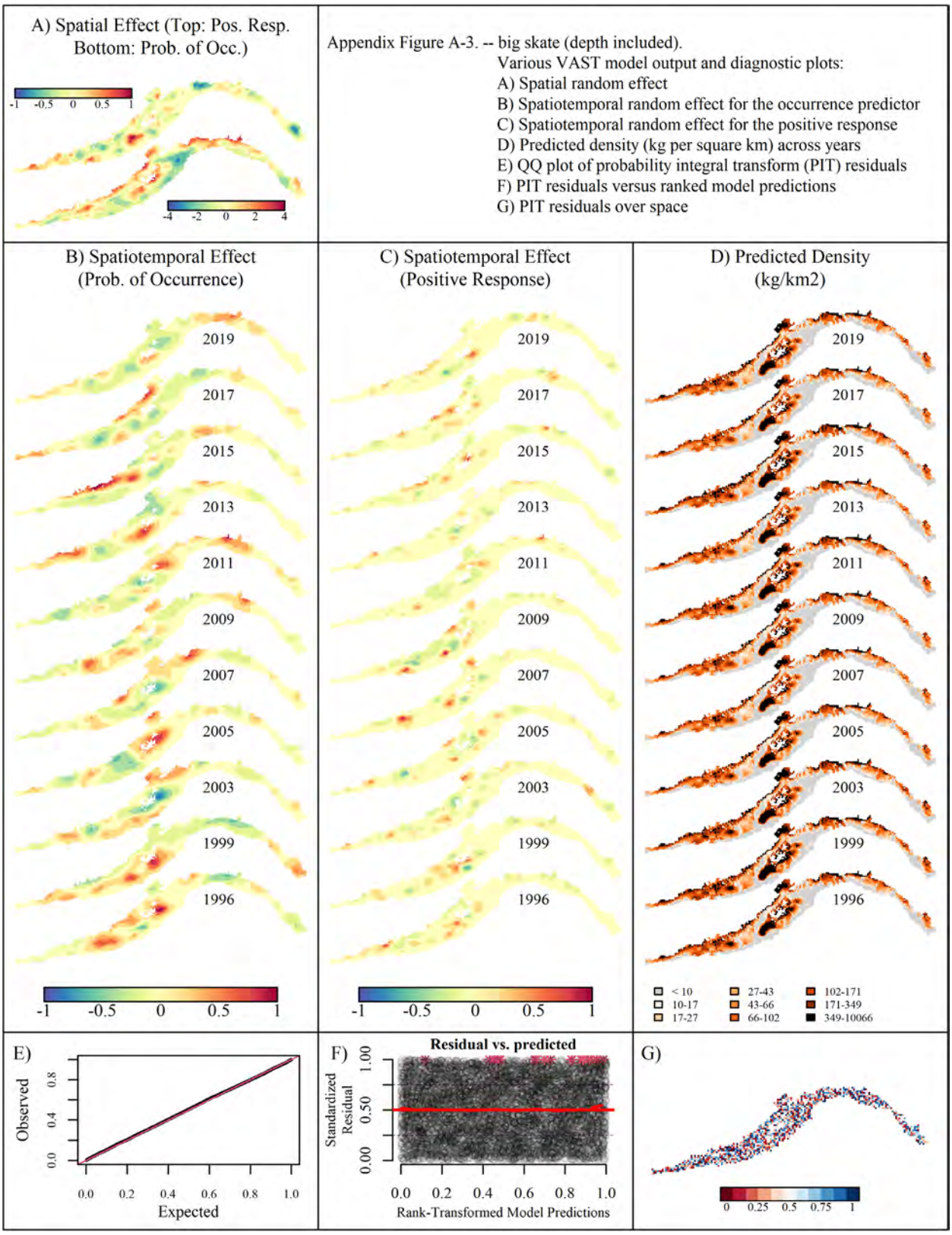
Appendix Figures A1-A26 show model outputs and diagnostic plots from a univariate vector autoregressive spatiotemporal (VAST) model. Some species have depth included as a quadratic linear effect noted in Table 3. All plots have the same layout: A) spatial random effect, B) spatiotemporal random effect on occurrence over time, C) spatiotemporal random effect on the positive component of the response over time, and D) predicted density over time. E) Probability integral transform (PIT) residual QQ plot, F) PIT residual versus ranked transformed model predictions, and G) PIT residuals over space.

Gulf of Alaska groundfishes comprised a diverse range of spatial distributions, as summarized in Appendix A. Pacific halibut (Appendix Fig. A-15), arrowtooth flounder (Appendix Fig. A-1), flathead sole (Appendix Fig. A-7), rex sole (Appendix Fig. A-17), Pacific cod (Appendix Fig. A-14), walleye pollock (Appendix Fig. A-25) and big and longnose skates (Appendix Figs. A-3 and A-11) were broadly geographically distributed and most prevalent along the continental shelf. Pacific ocean perch (Appendix Fig. A-16), harlequin rockfish (Appendix Fig. A-10), and dusky rockfish (Appendix Fig. A-6) were distributed primarily in deeper depths, along the outer continental shelf and upper slope, and often with ephemeral hotspots of high density. Shortraker rockfish (Appendix Fig. A-20), sablefish (Appendix Fig. A-18),), and shortspine thornyhead (Appendix Fig. A-21) were most prevalent on the continental slope, with less variability in spatial distribution over time than many of the species found more

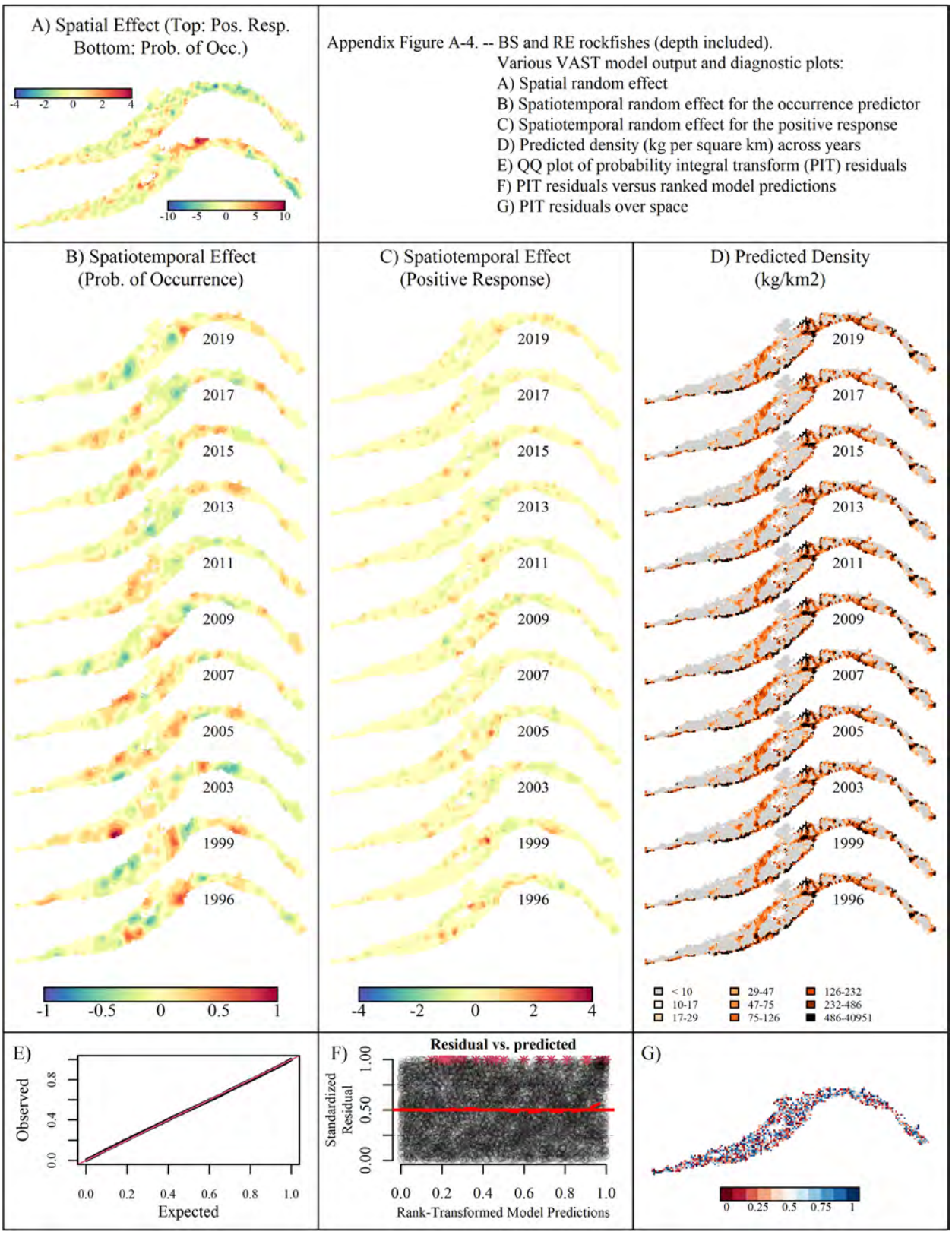
commonly on the continental shelf. A handful of species were restricted to certain parts of the spatial domain. Northern rockfish (Appendix Fig. A-13) were most prevalent along the outer continental shelf and upper slope in the western area of the domain. Dover sole (Appendix Fig. A-5) were prevalent along the continental slope and in deep gullies everywhere except the far western portion of the domain. The rock soles were distributed shallower and were more shelf-associated than Dover sole and northern rockfish, with northern rock sole (Appendix Fig. A-12) restricted to the western and central portions of the spatial domain, and southern rock sole (Appendix Fig. A-23) was distributed similarly with the exception of some isolated hotspots in the Southeast GOA. Pacific spiny dogfish (Appendix Fig. A-24) were found at shallow to moderate depths and silvergray rockfish (Appendix Fig. A-22) at deeper depths across their distributions, which were restricted to the central and eastern GOA. Conversely, Atka mackerel (Appendix Fig. A-2) and giant octopus (Appendix Fig. A-9) were restricted to the western/central part of the spatial domain. Visual evaluations of the PIT residuals showed no aberrant departures from model assumptions.


A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-4 -2 0 2 4
Appendix Figure A-1. -- arrowtooth flounder (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-4 -2 0 2 4
< 10
10-379
379-926
926-1562
1562-2413
2413-3653
3653-5600
5600-9508
9508-507892
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

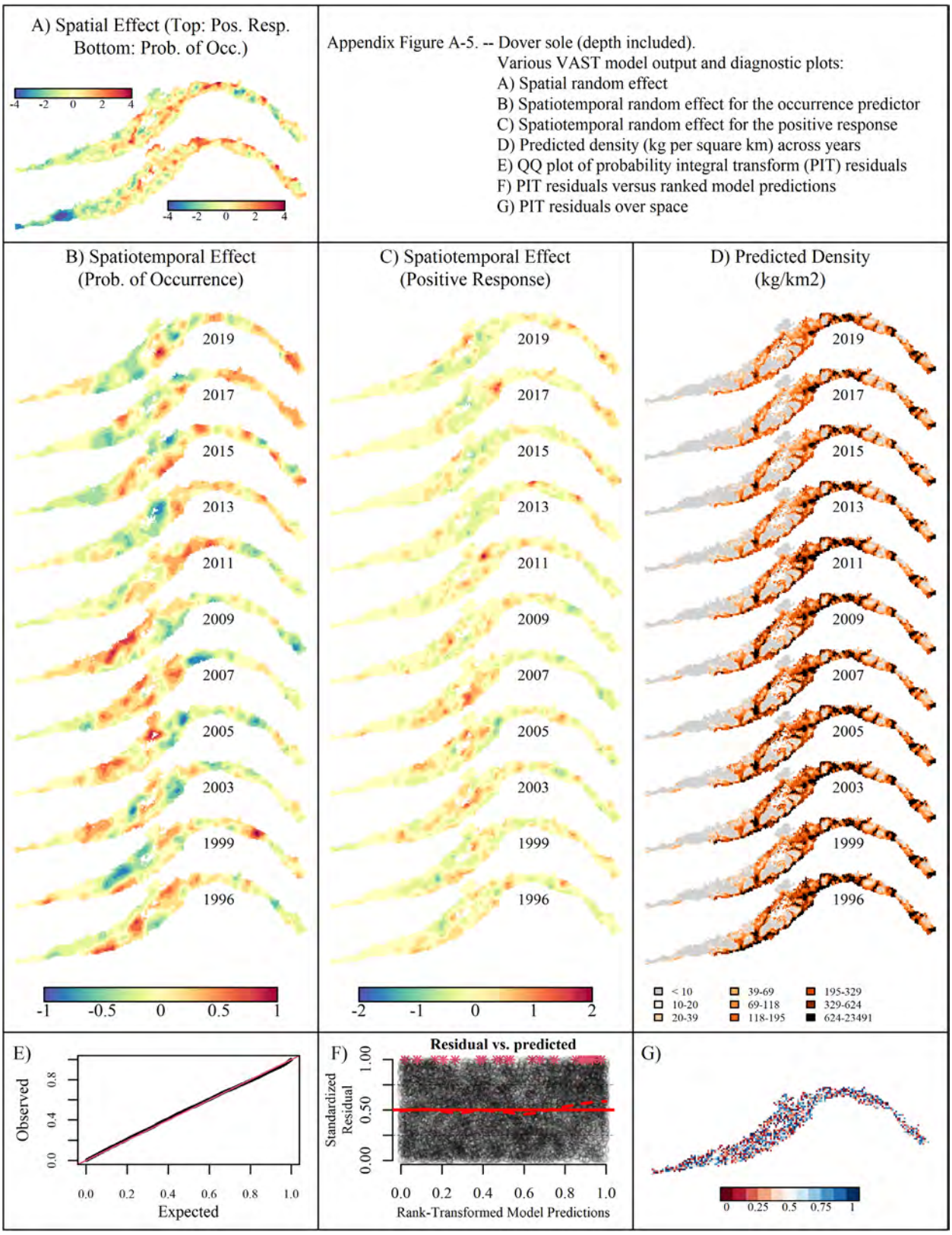

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-2 -1 0 1 2
-5 0 5
Appendix Figure A-2. -- Atka mackerel (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-4 -2 0 2 4
-5 0 5
< 10
10-13
13-17
17-24
24-35
35-54
54-91
91-220
220-99809
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1



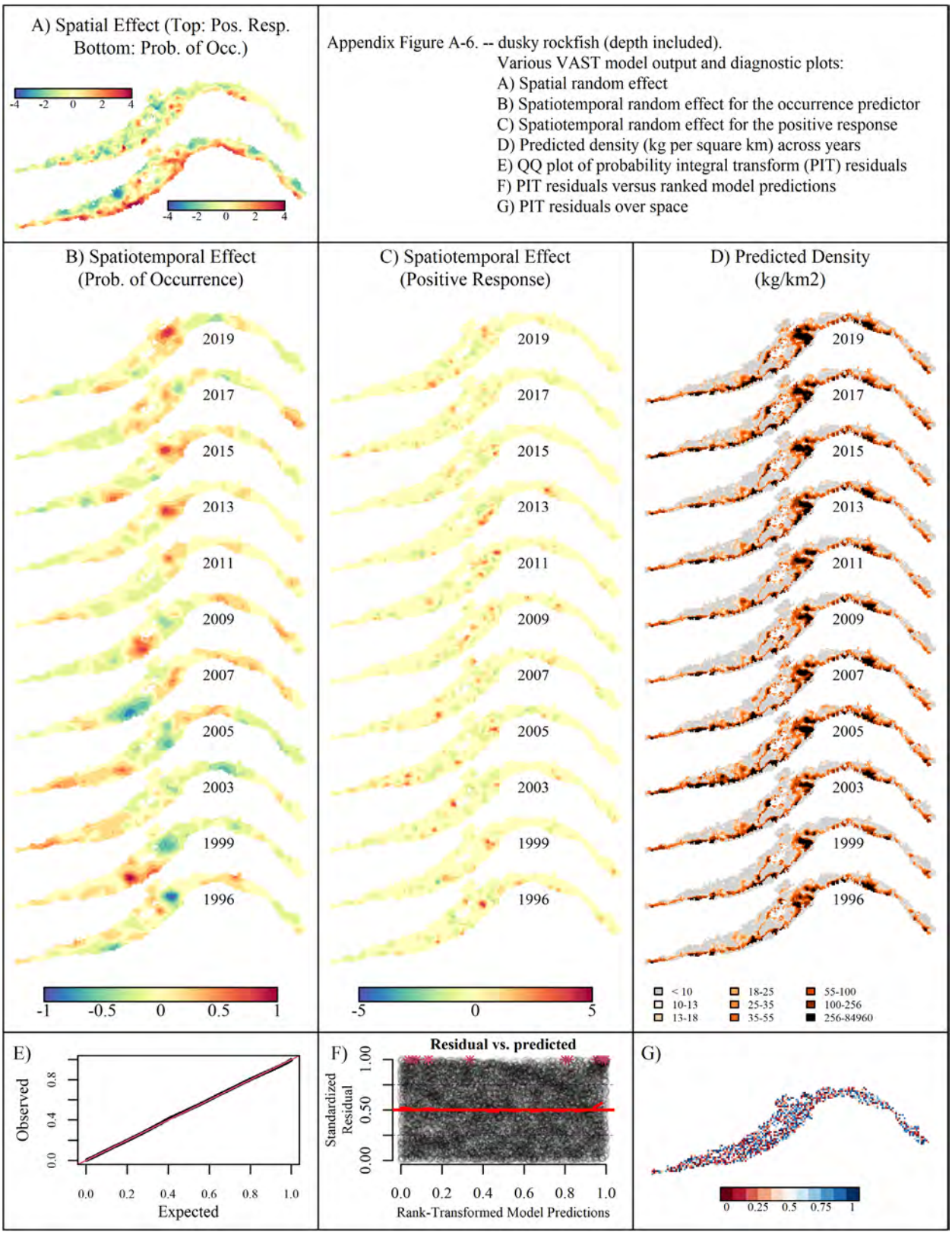
Appendix Figure A-3. -- big skate (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

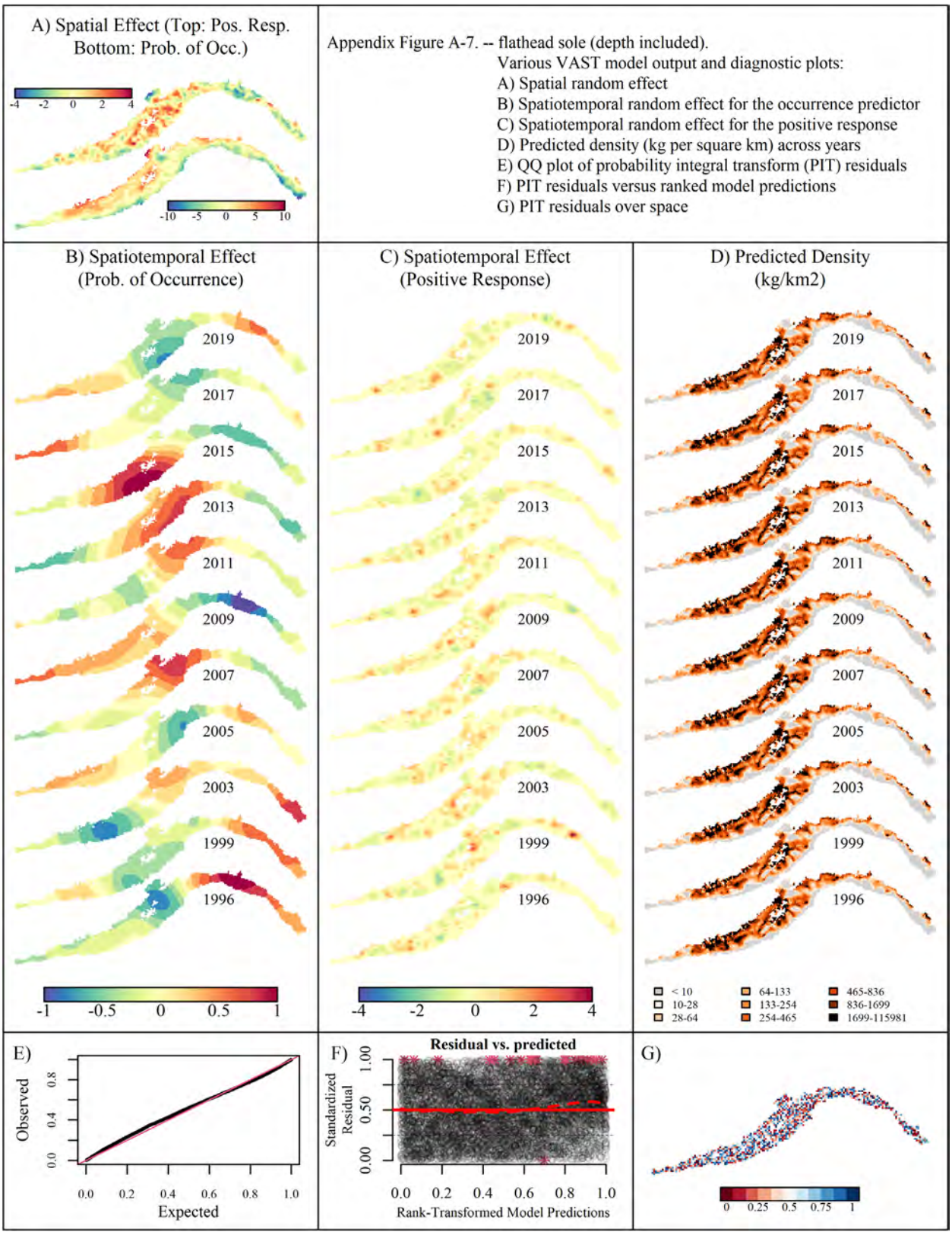


Appendix Figure A-4. -- BS and RE rockfishes (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

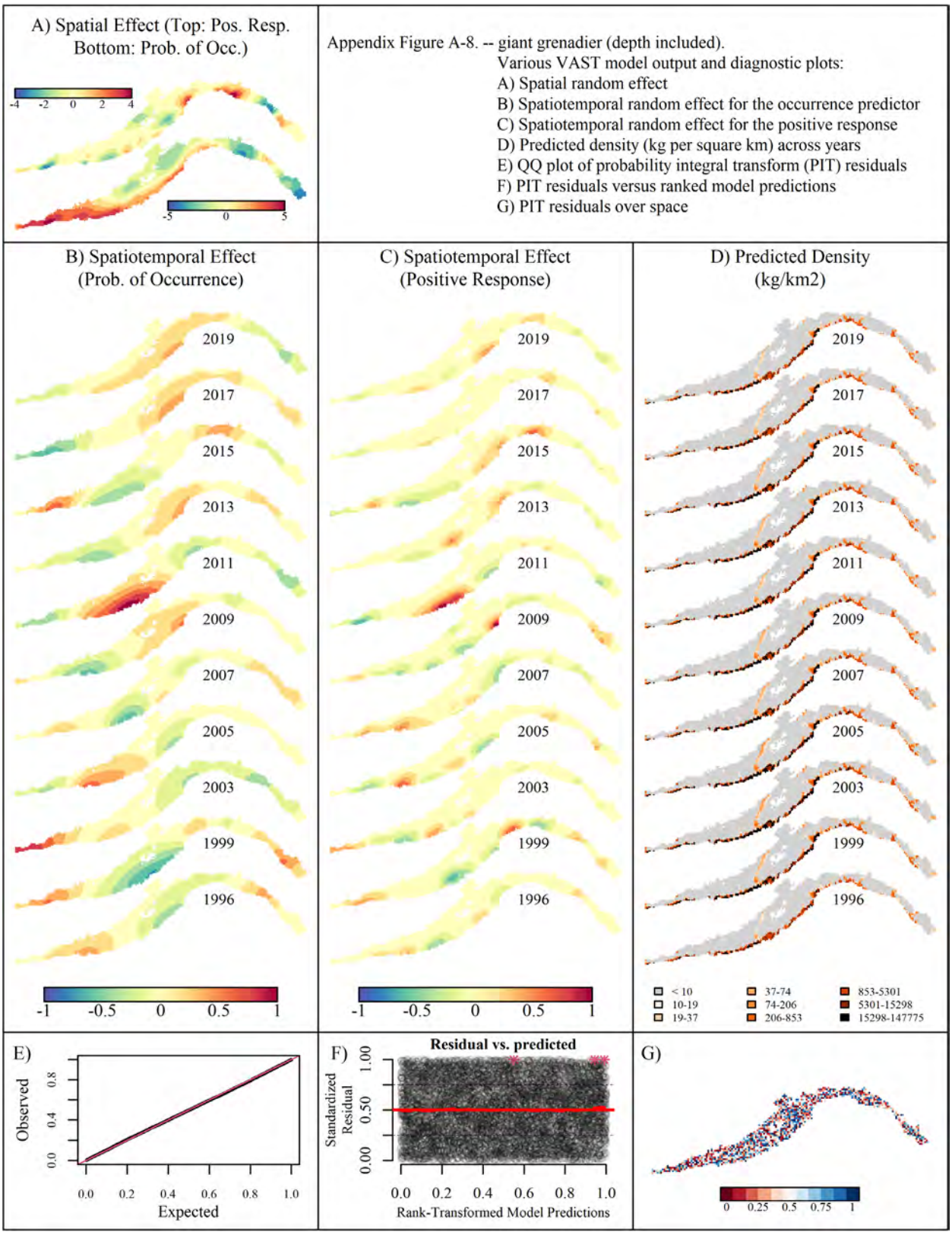


Appendix Figure A-5. -- Dover sole (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

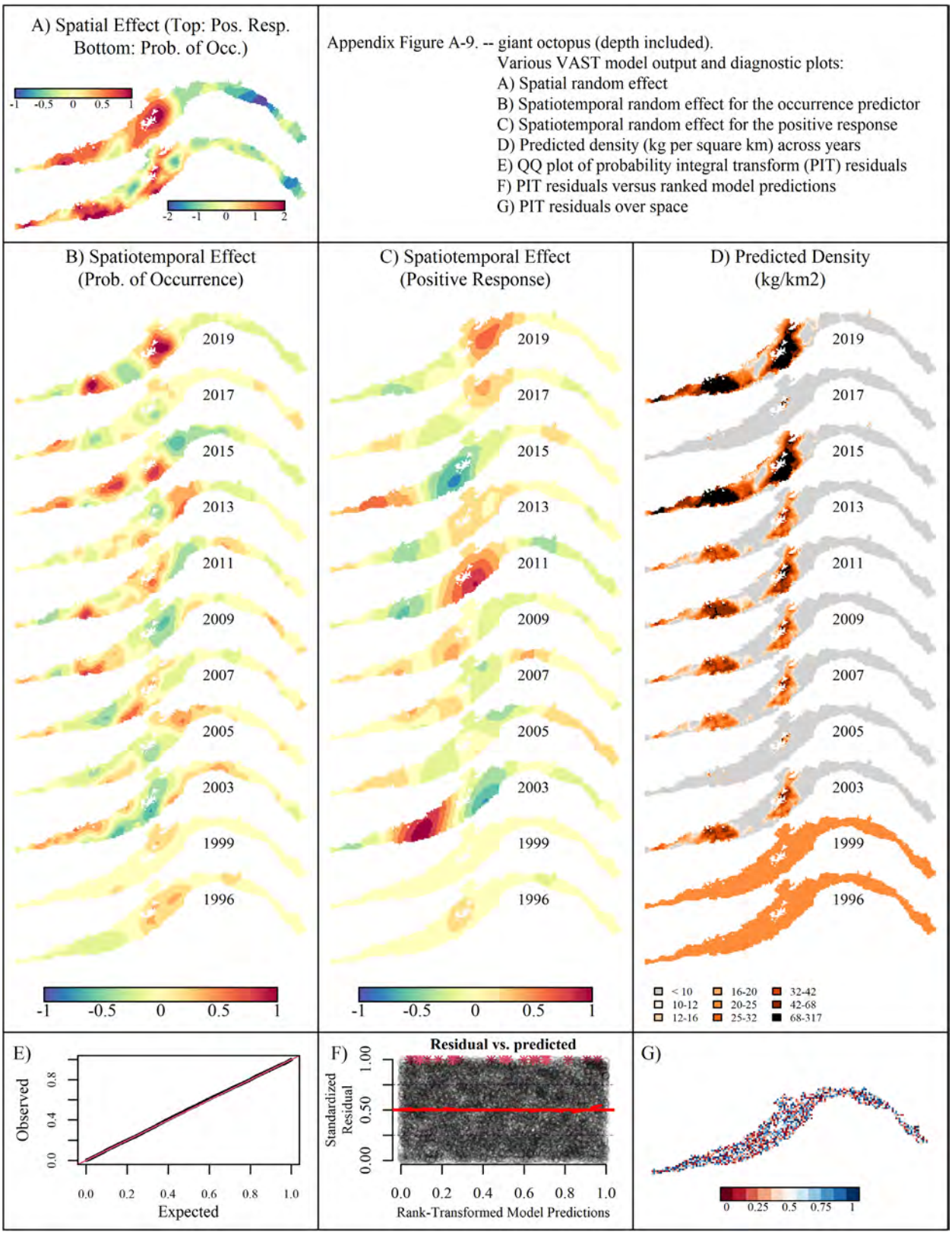

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-4 -2 0 2 4
Appendix Figure A-6. -- dusky rockfish (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-5 0 5
< 10
10-13
13-18
18-25
25-35
35-55
55-100
100-256
256-84960
E)
Observed
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0
Expected
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
0.0 0.2 0.4 0.6 0.8 1.0
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1



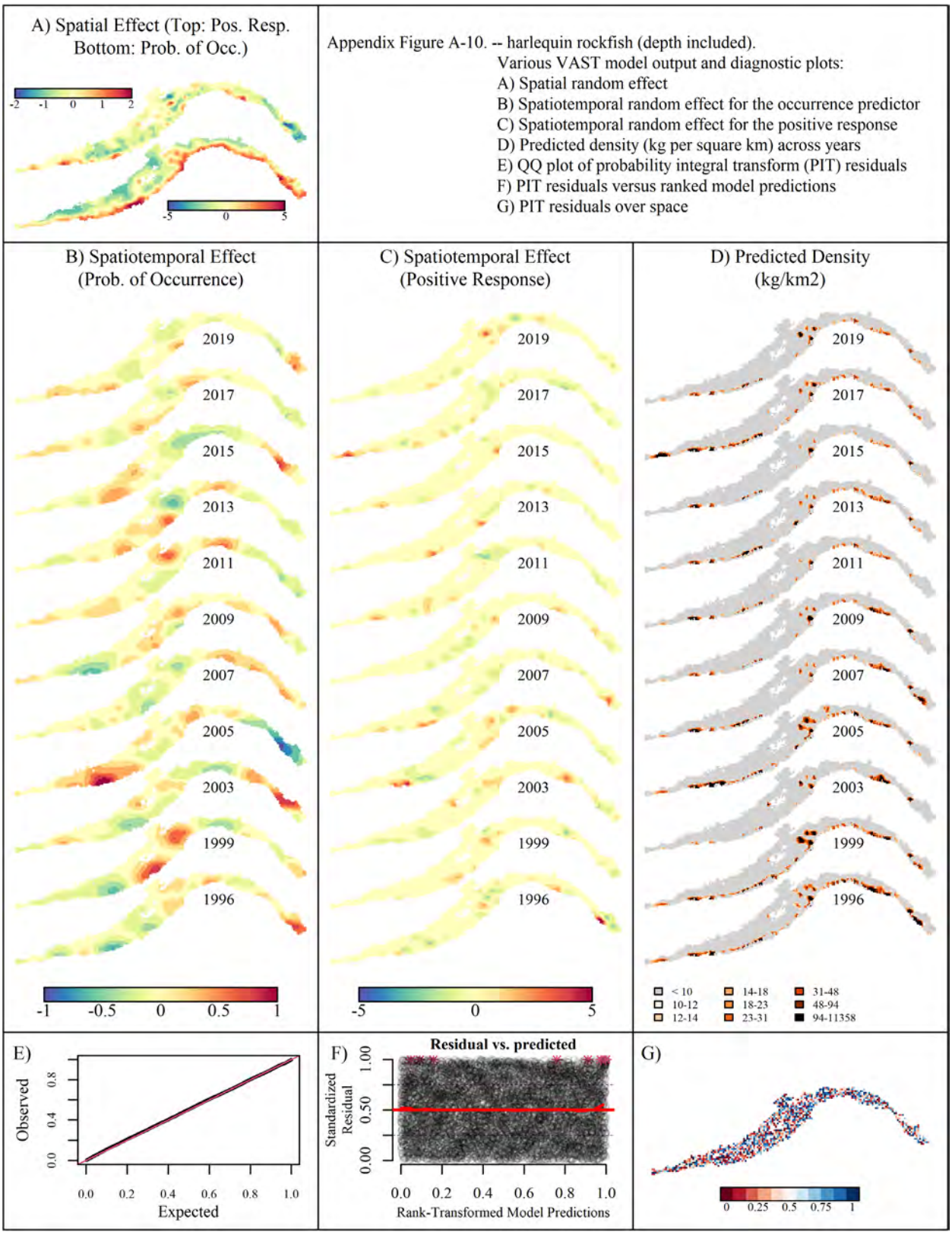
Appendix Figure A-7. -- flathead sole (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space



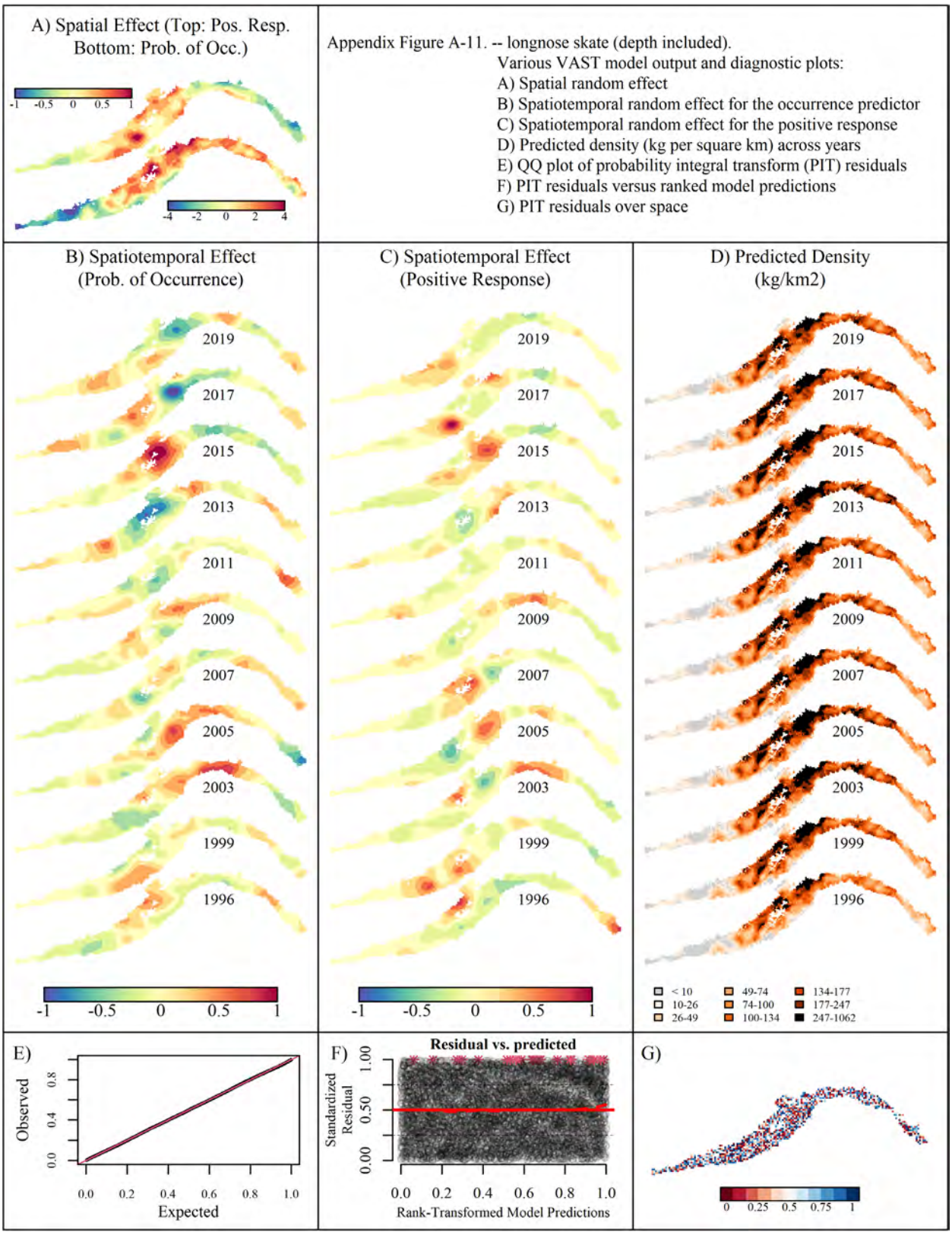
Appendix Figure A-8. -- giant grenadier (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space



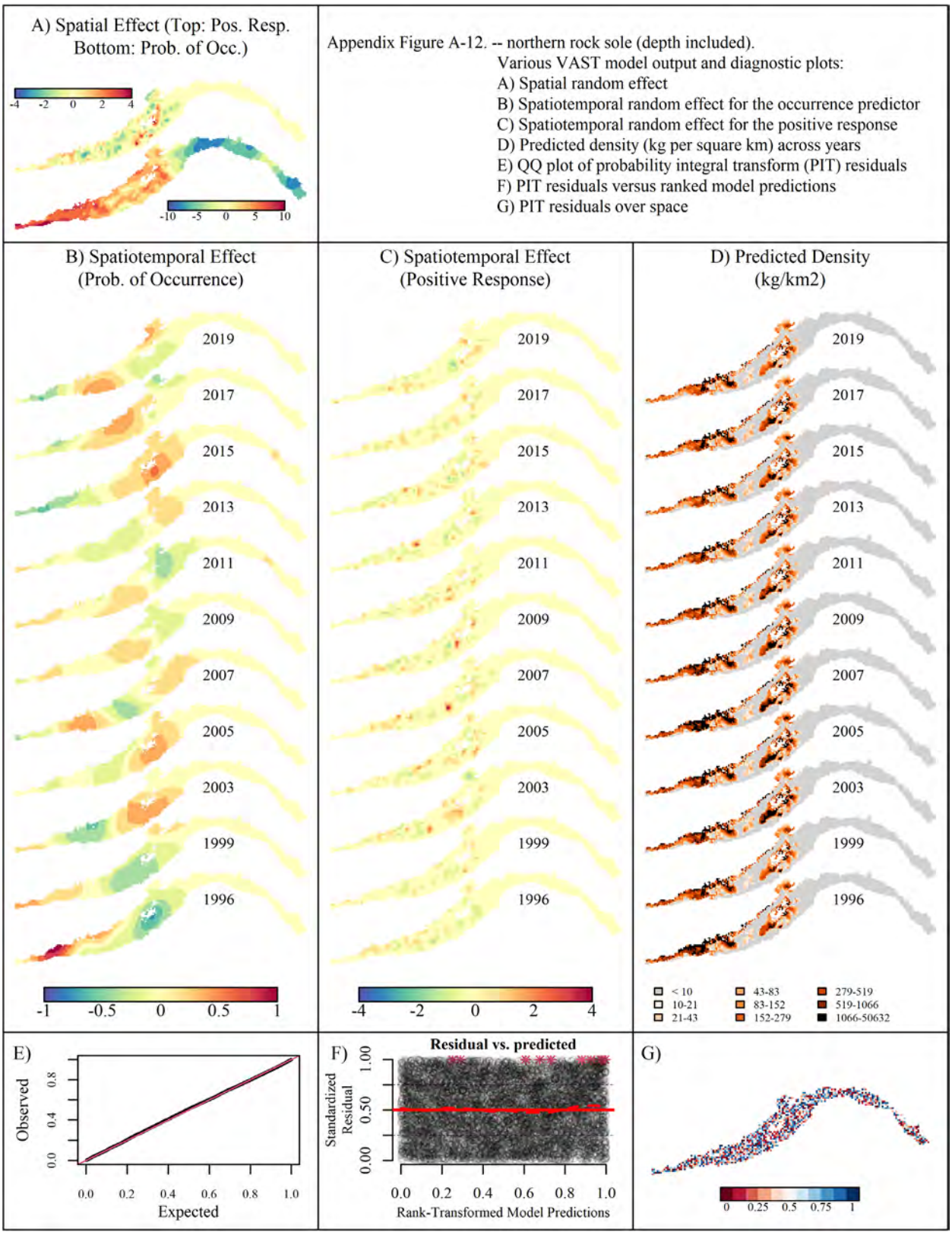
Appendix Figure A-9. -- giant octopus (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

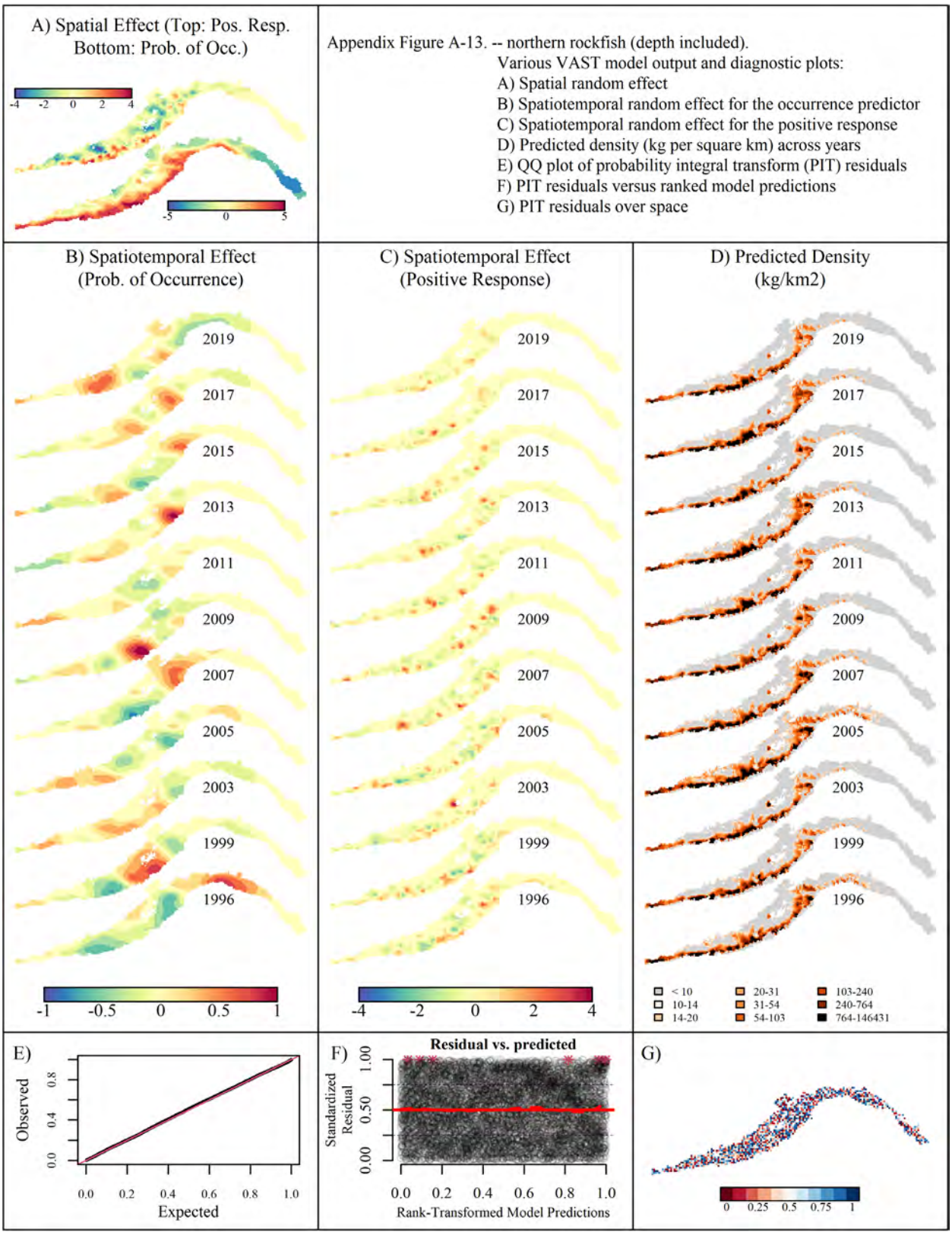


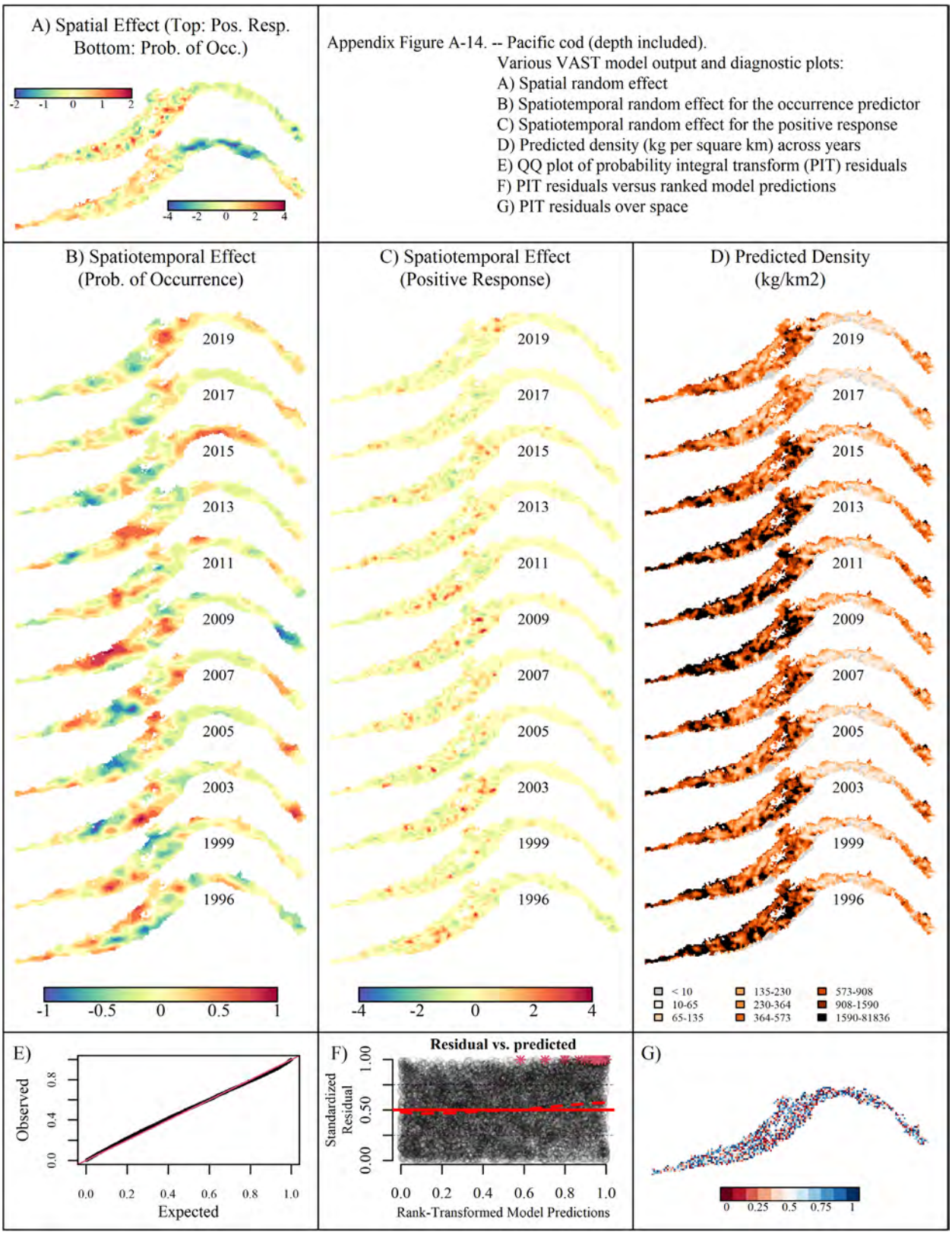
Appendix Figure A-10. -- harlequin rockfish (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

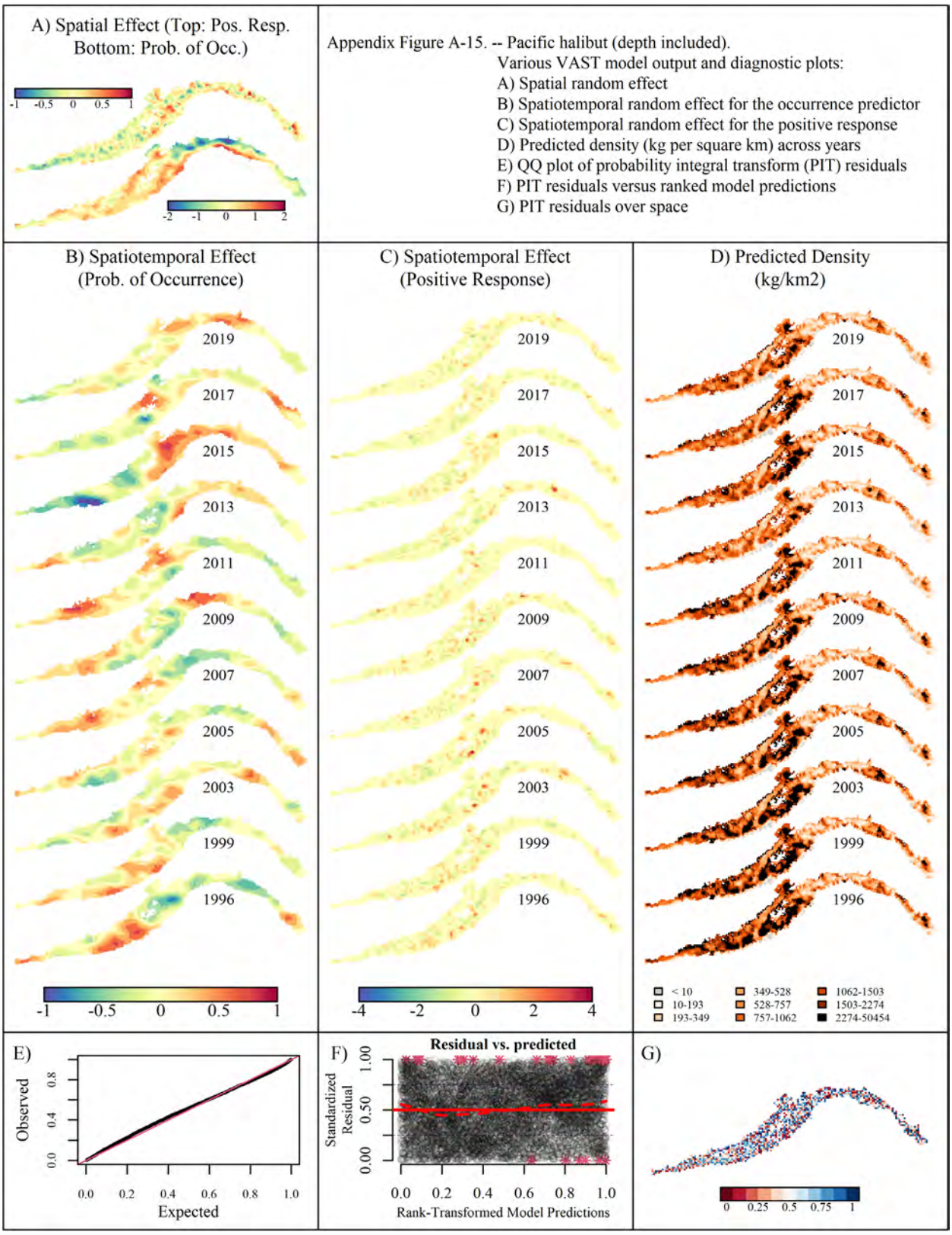


Appendix Figure A-11. -- longnose skate (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

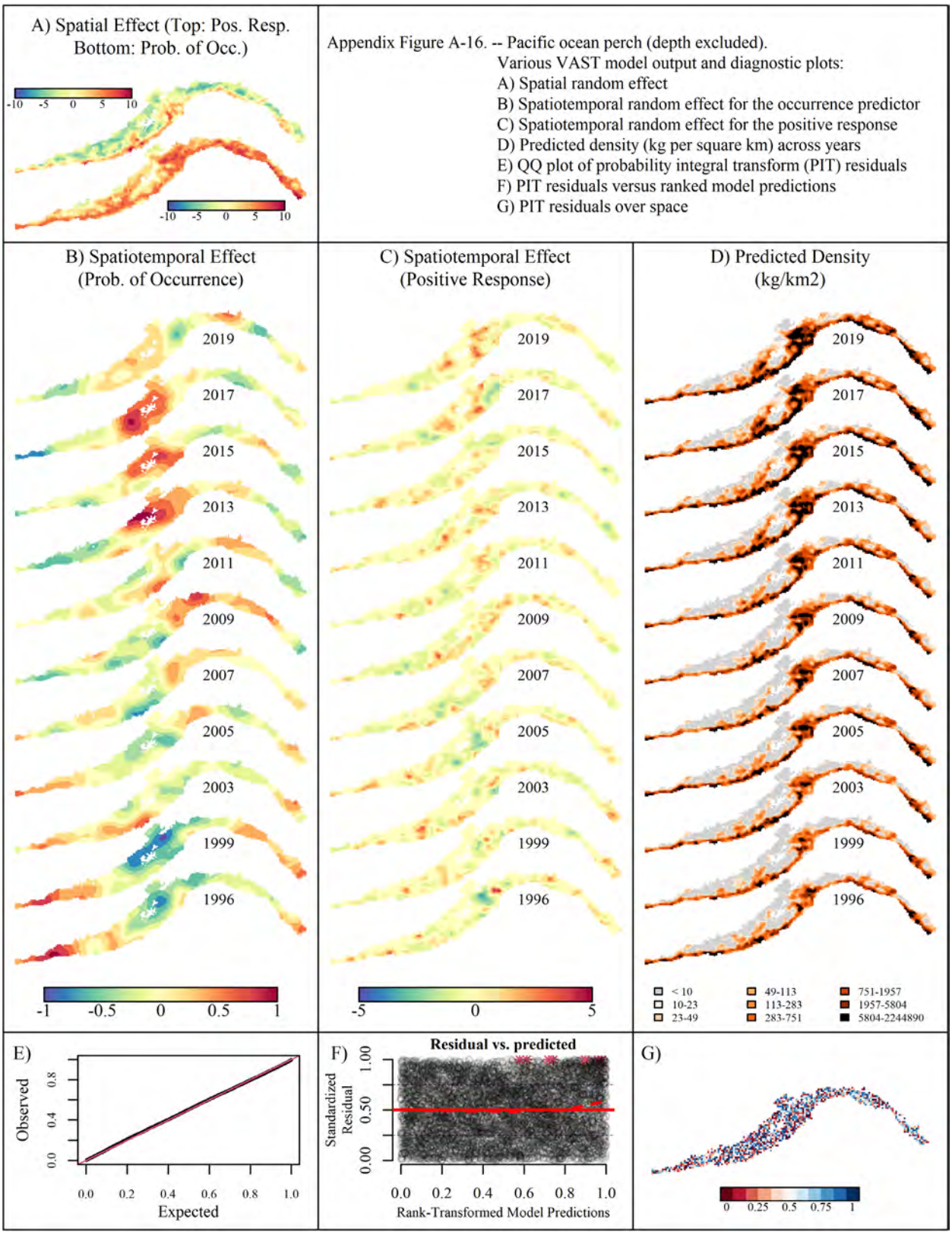

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-10 -5 0 5 10
Appendix Figure A-12. -- northern rock sole (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-4 -2 0 2 4
< 10
10-21
21-43
43-83
83-152
152-279
279-519
519-1066
1066-50632
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

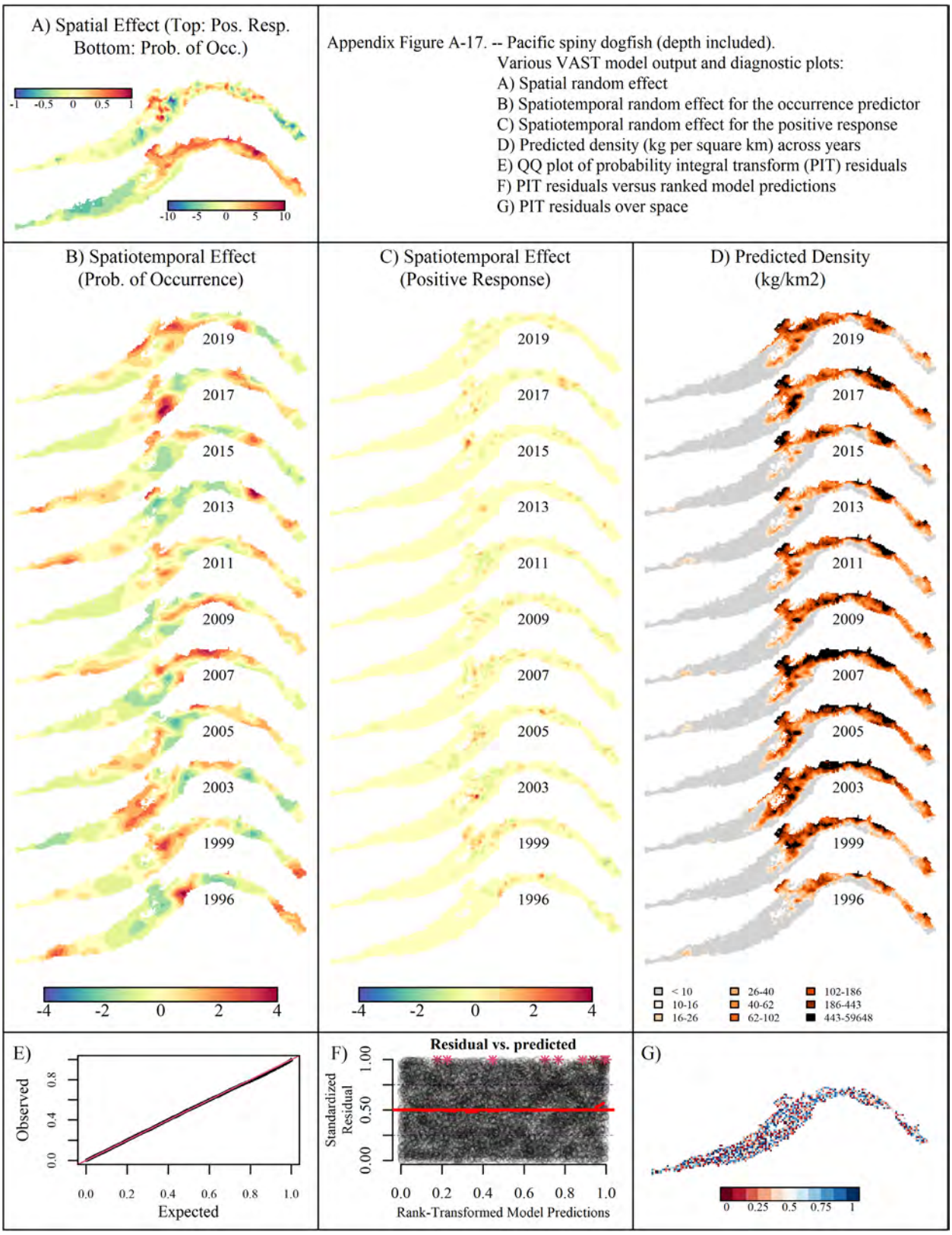

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-5 0 5
Appendix Figure A-13. -- northern rockfish (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-4 -2 0 2 4
< 10
10-14
14-20
20-31
31-54
54-103
103-240
240-764
764-146431
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1


A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-2 -1 0 1 2
-4 -2 0 2 4
Appendix Figure A-14. -- Pacific cod (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-4 -2 0 2 4
< 10
10-65
65-135
135-230
230-364
364-573
573-908
908-1590
1590-81836
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

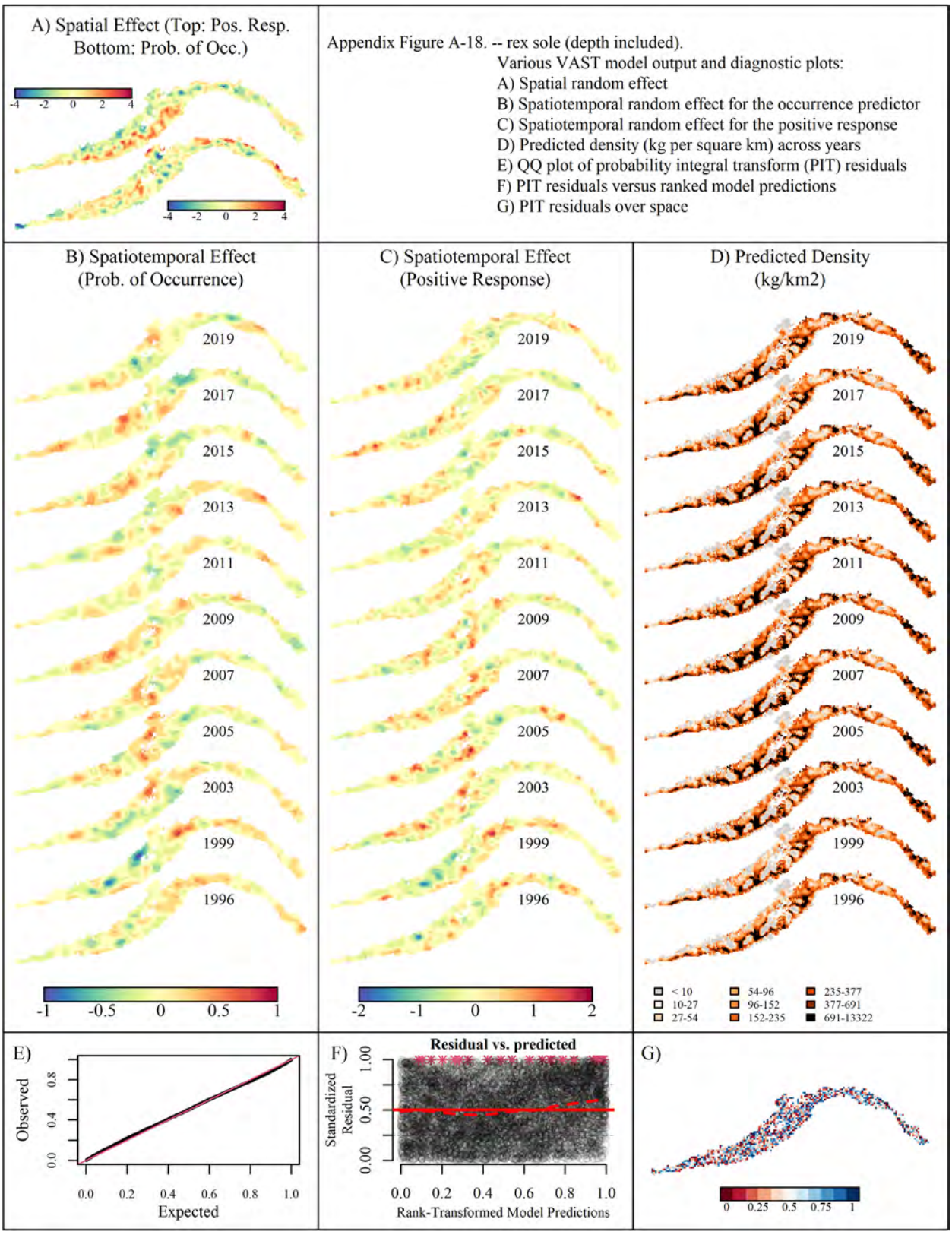


Appendix Figure A-15. -- Pacific halibut (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

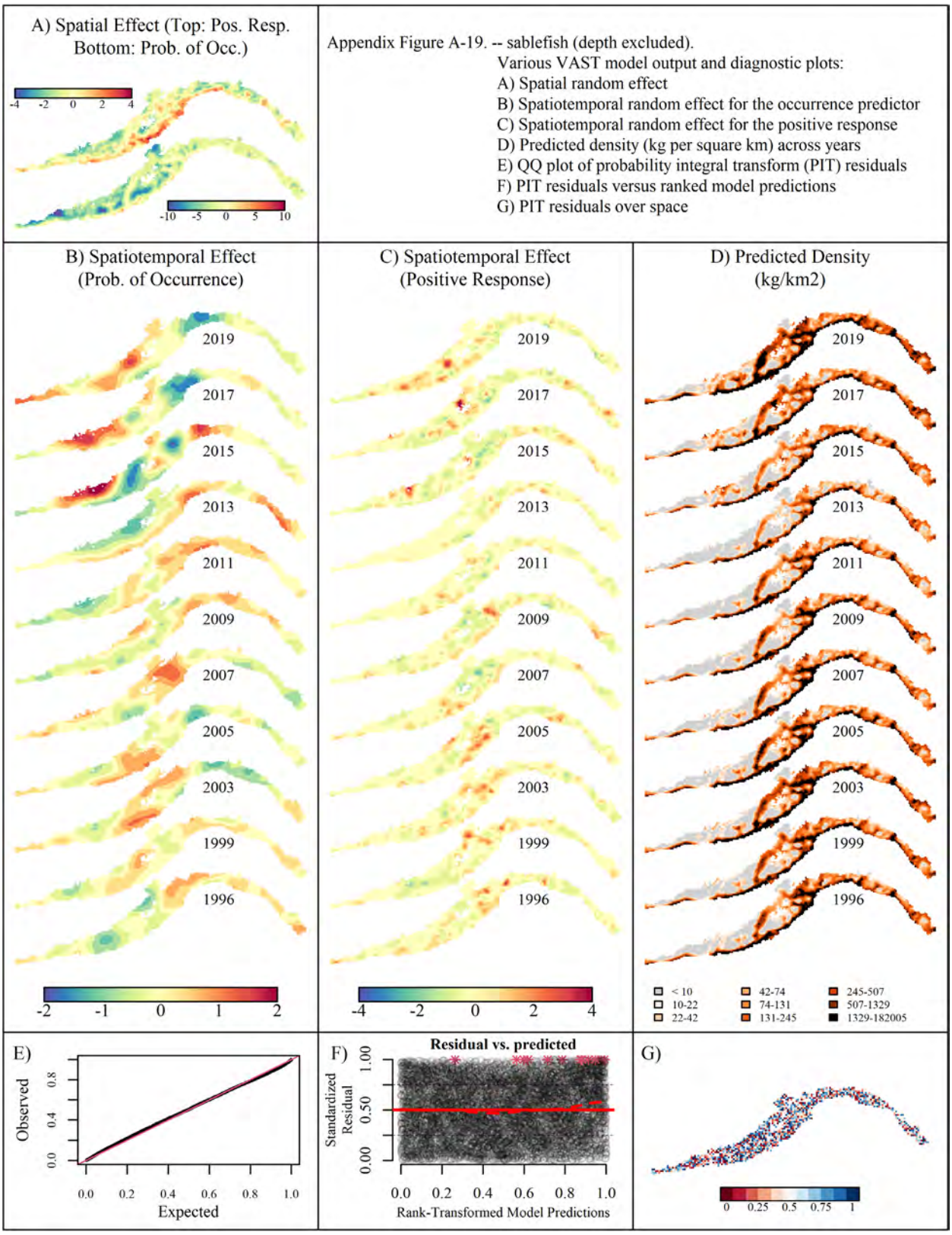

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-10 -5 0 5 10
-10 -5 0 5 10
Appendix Figure A-16. -- Pacific ocean perch (depth excluded).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-5 0 5
< 10
10-23
23-49
49-113
113-283
283-751
751-1957
1957-5804
5804-2244890
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

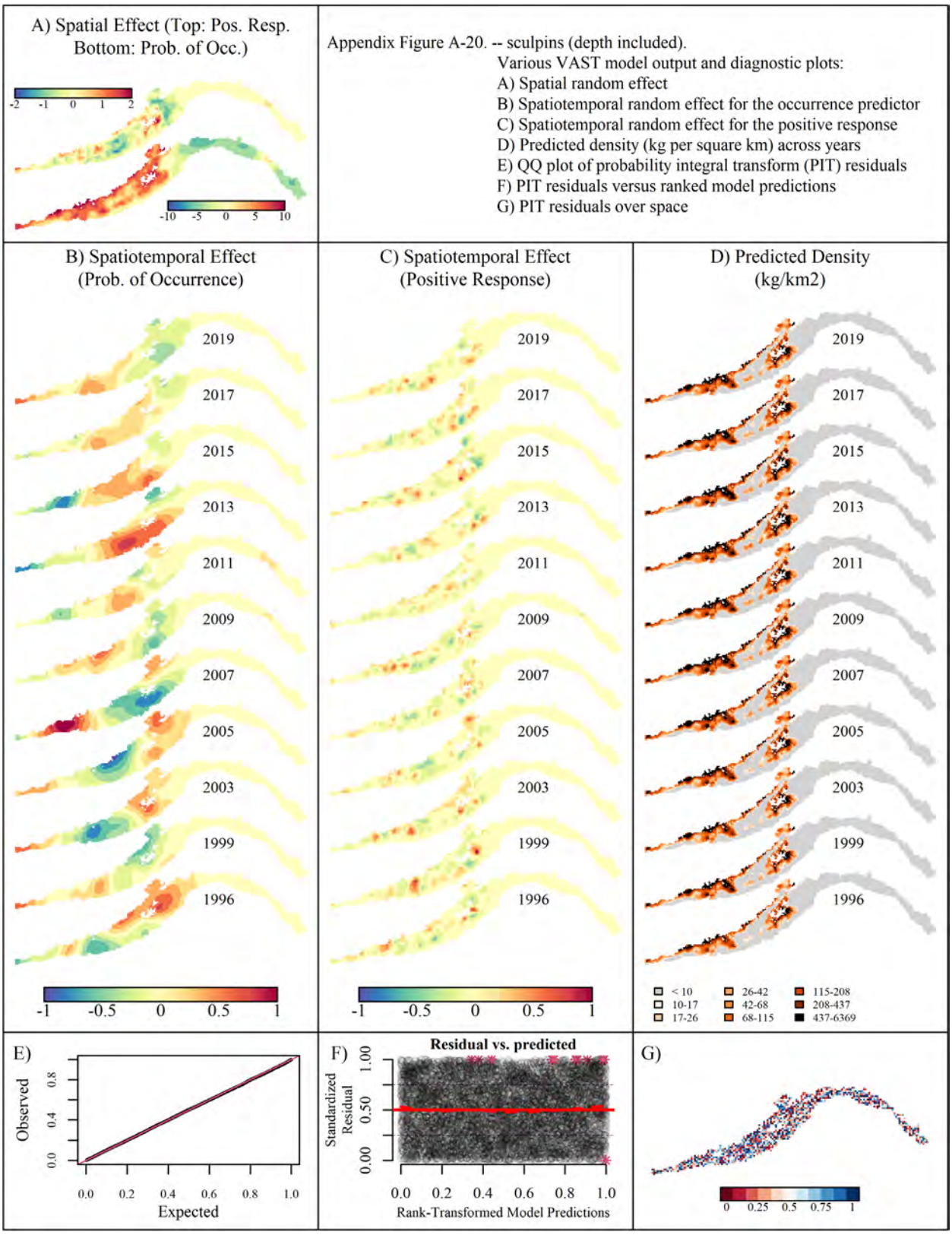


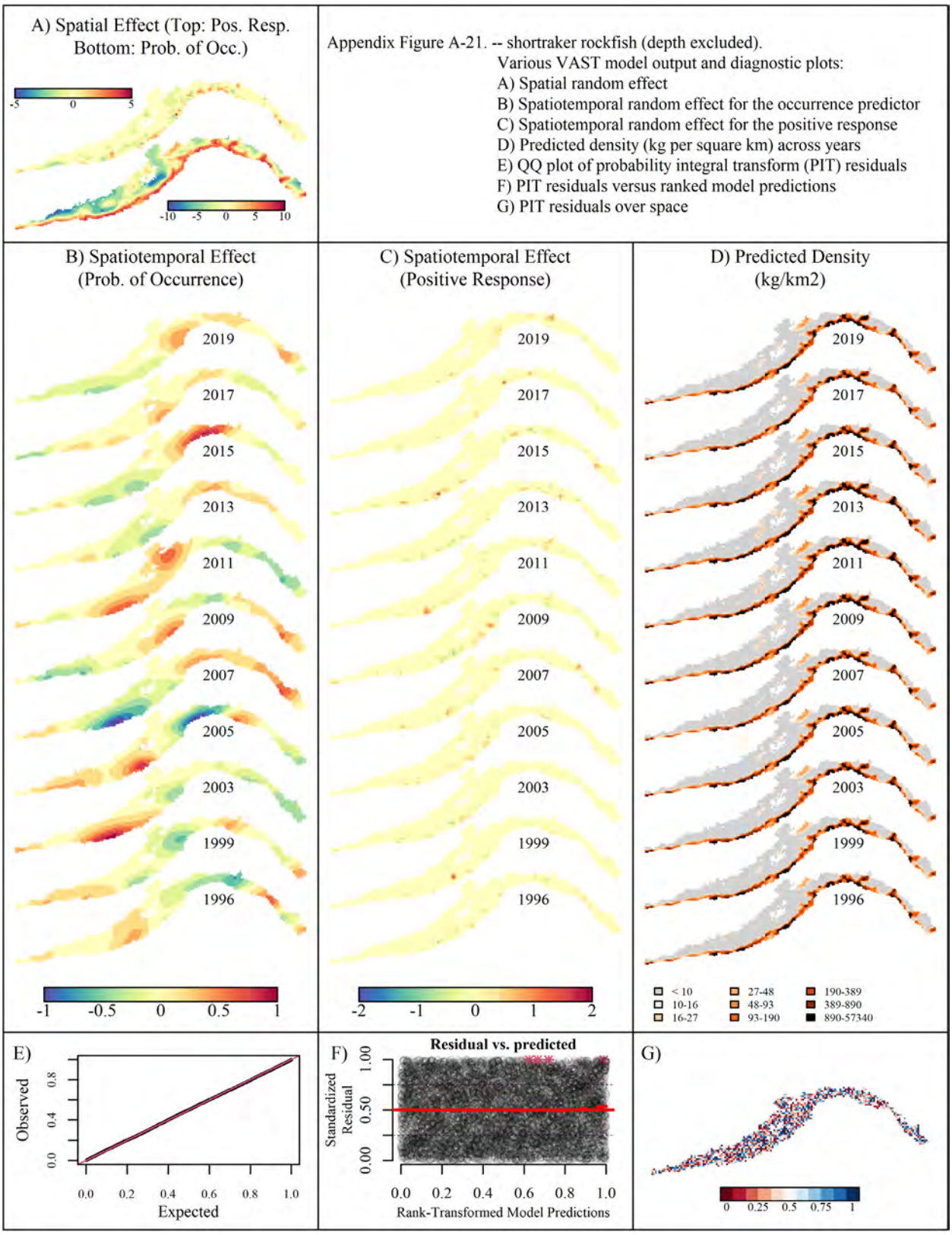
Appendix Figure A-17. -- Pacific spiny dogfish (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

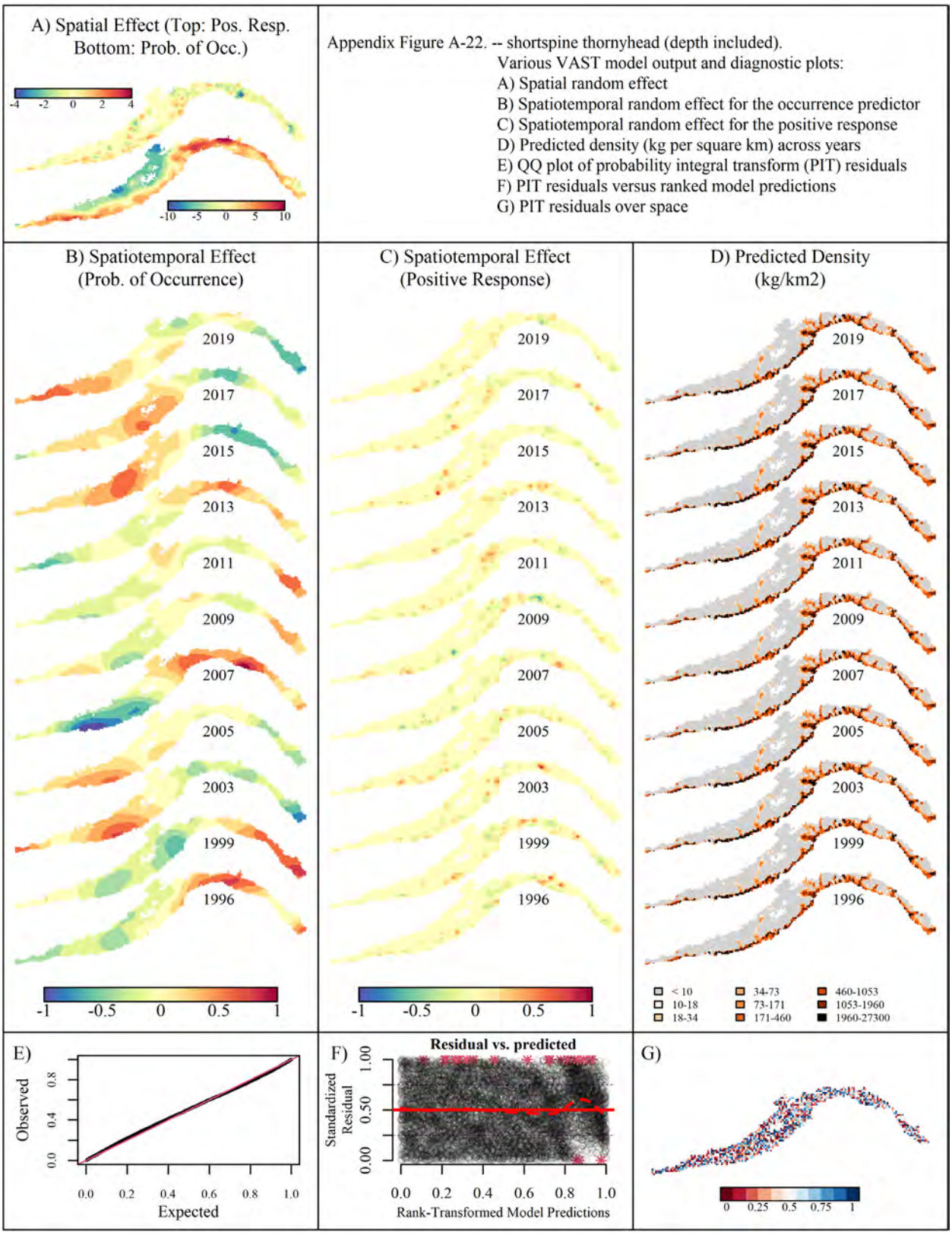


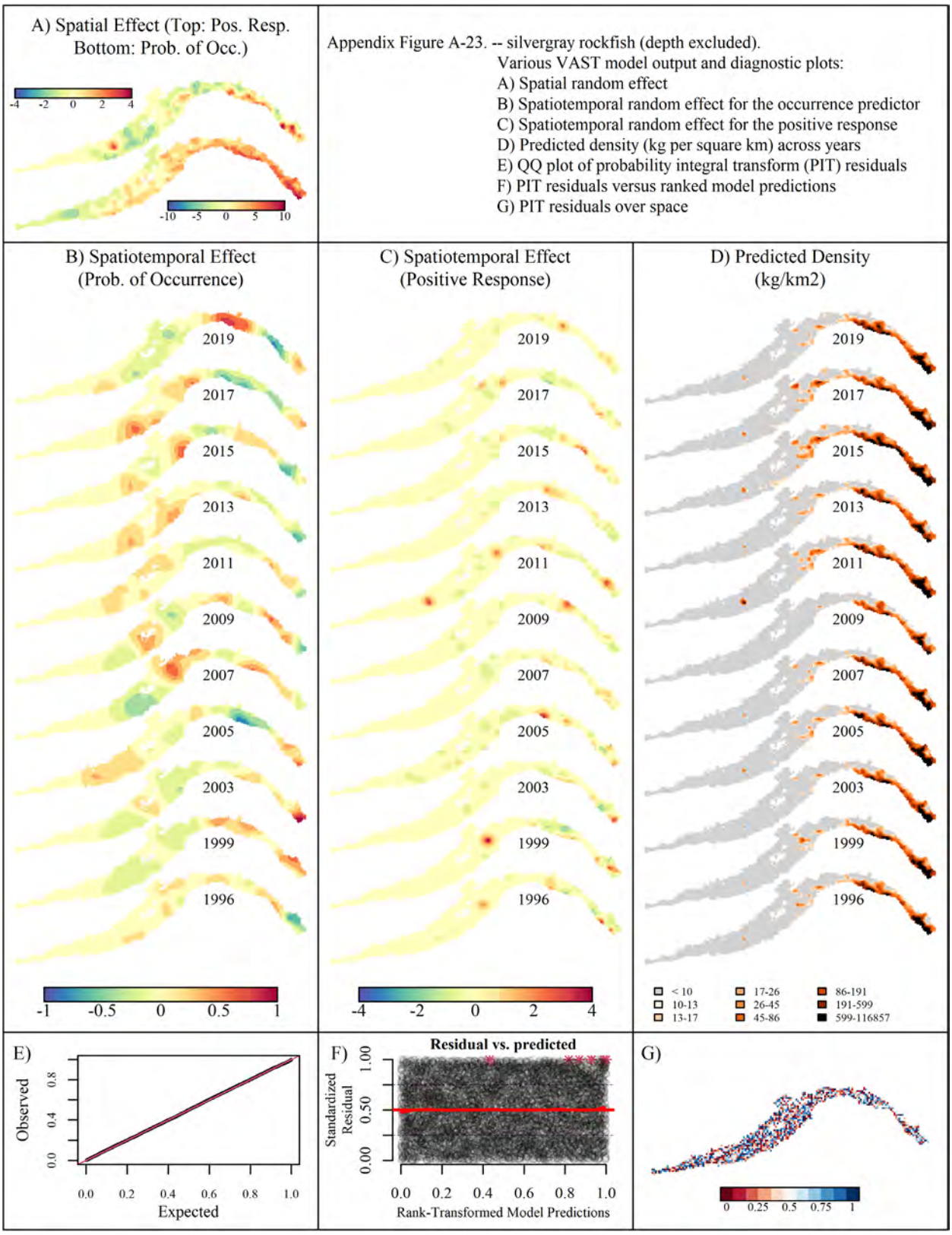
Appendix Figure A-18. -- rex sole (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

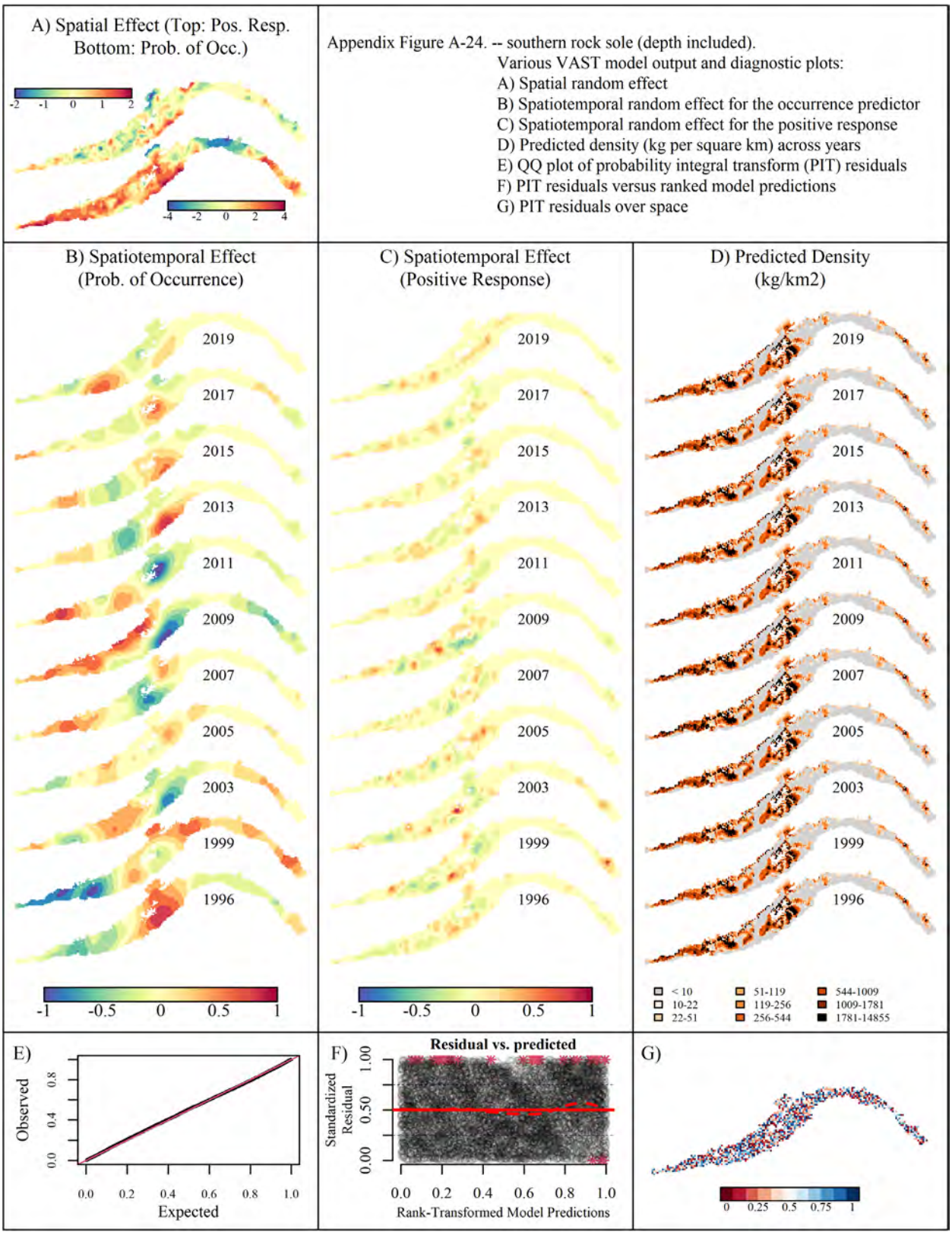

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-10 -5 0 5 10
Appendix Figure A-19. -- sablefish (depth excluded).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-2 -1 0 1 2
-4 -2 0 2 4
< 10
10-22
22-42
42-74
74-131
131-245
245-507
507-1329
1329-182005
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

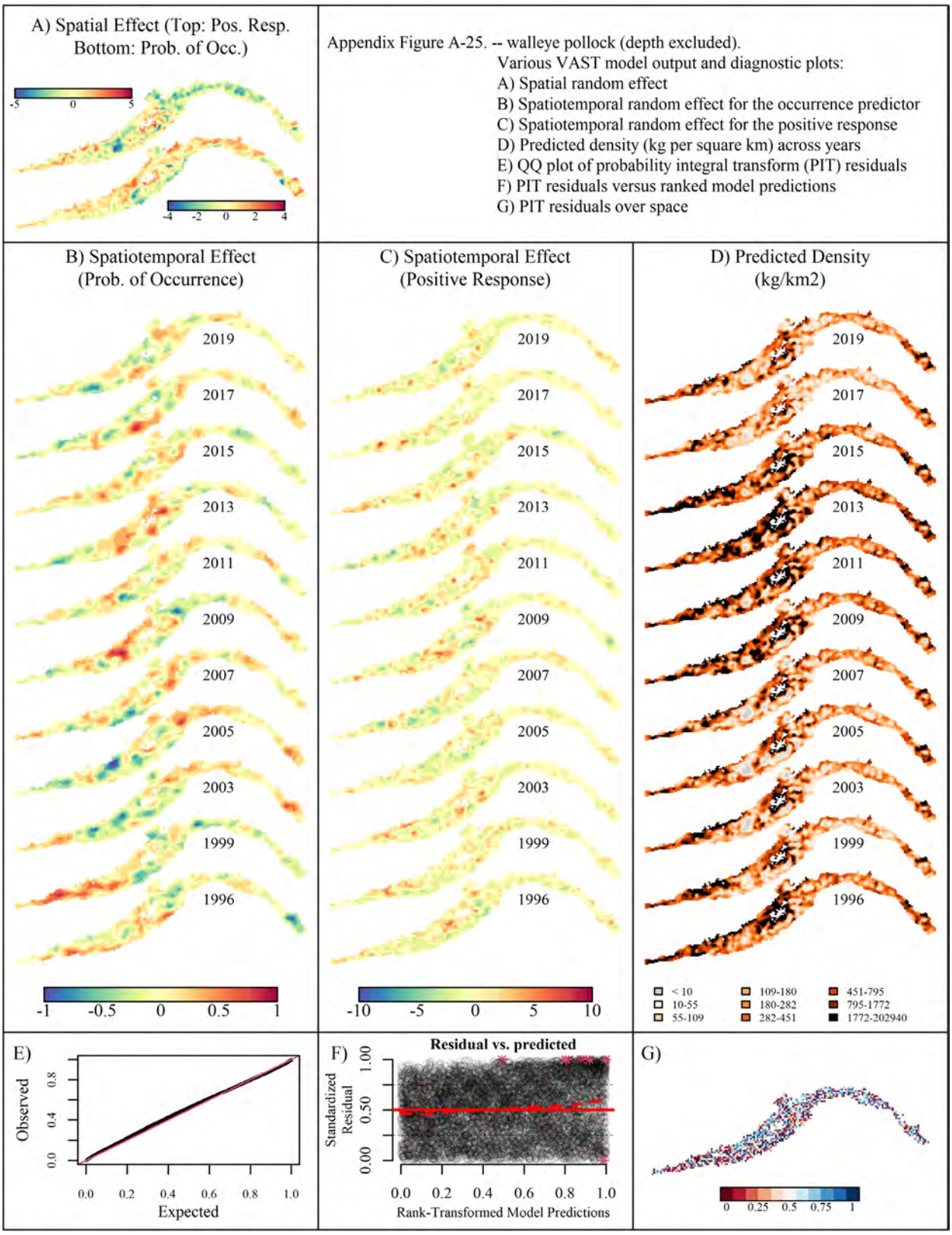

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-2 -1 0 1 2
-10 -5 0 5 10
Appendix Figure A-20. -- sculpins (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
< 10
10-17
17-26
26-42
42-68
68-115
115-208
208-437
437-6369
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

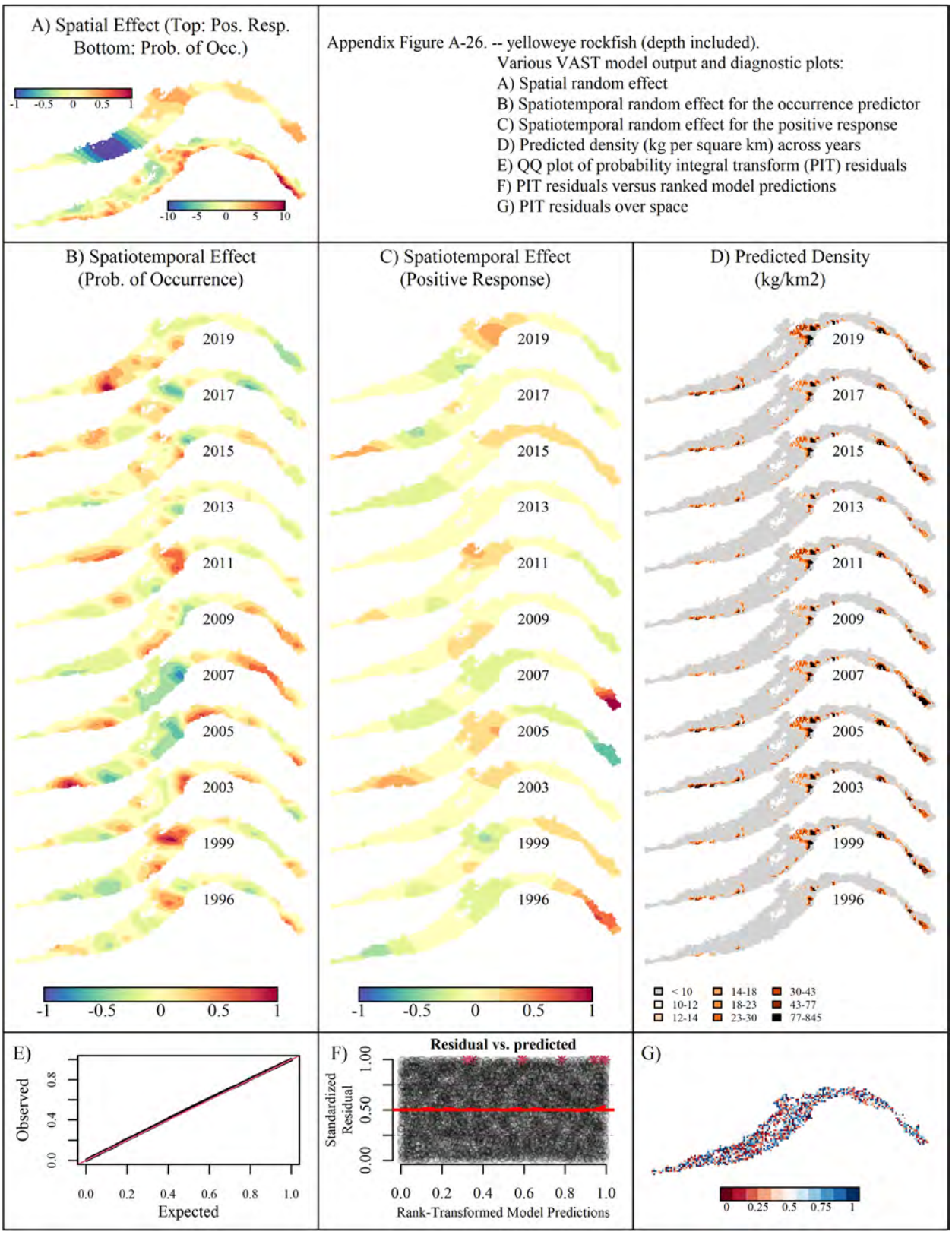

A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-5 0 5
-10 -5 0 5 10
Appendix Figure A-21. -- shortraker rockfish (depth excluded).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-2 -1 0 1 2
< 10
10-16
16-27
27-48
48-93
93-190
190-389
389-890
890-57340
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1


A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-10 -5 0 5 10
Appendix Figure A-22. -- shortspine thornyhead (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
< 10
10-18
18-34
34-73
73-171
171-460
460-1053
1053-1960
1960-27300
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1


A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-4 -2 0 2 4
-10 -5 0 5 10
Appendix Figure A-23. -- silvergray rockfish (depth excluded).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-4 -2 0 2 4
< 10
10-13
13-17
17-26
26-45
45-86
86-191
191-599
599-116857
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1


A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-2 -1 0 1 2
-4 -2 0 2 4
Appendix Figure A-24. -- southern rock sole (depth included).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
< 10
10-22
22-51
51-119
119-256
256-544
544-1009
1009-1781
1781-14855
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1


A) Spatial Effect (Top: Pos. Resp. Bottom: Prob. of Occ.)
-5 0 5
-4 -2 0 2 4
Appendix Figure A-25. -- walleye pollock (depth excluded).
Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space
B) Spatiotemporal Effect (Prob. of Occurrence)
C) Spatiotemporal Effect (Positive Response)
D) Predicted Density (kg/km2)
2019
2017
2015
2013
2011
2009
2007
2005
2003
1999
1996
-1 -0.5 0 0.5 1
-10 -5 0 5 10
< 10
10-55
55-109
109-180
180-282
282-451
451-795
795-1772
1772-202940
E)
Observed
Expected
0.0 0.2 0.4 0.6 0.8 1.0
F)
Residual vs. predicted
Standardized Residual
0.00 0.50 1.00
Rank-Transformed Model Predictions
G)
0 0.25 0.5 0.75 1

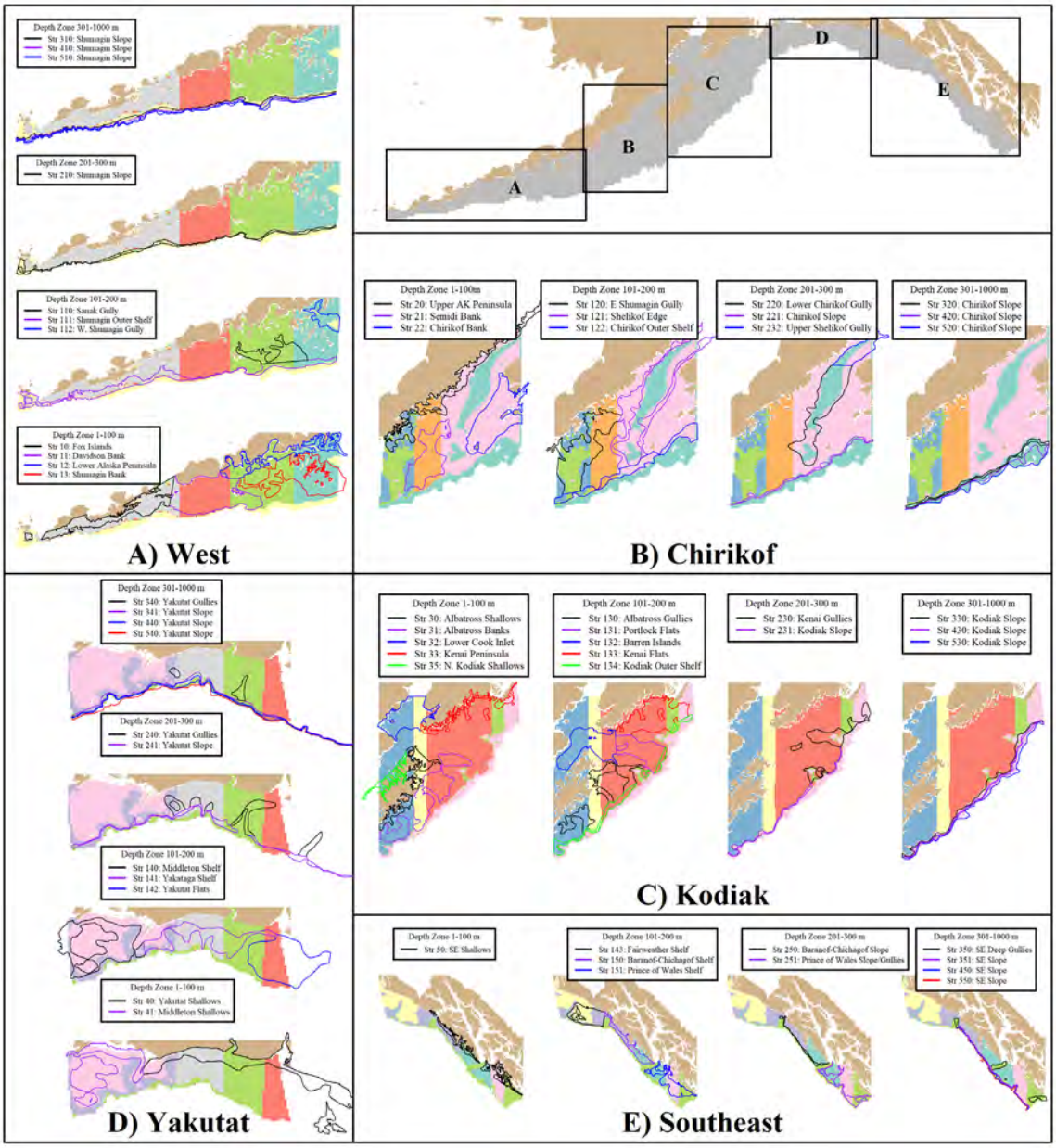


Appendix Figure A-26. -- yelloweye rockfish (depth included). Various VAST model output and diagnostic plots:
A) Spatial random effect
B) Spatiotemporal random effect for the occurrence predictor
C) Spatiotemporal random effect for the positive response
D) Predicted density (kg per square km) across years
E) QQ plot of probability integral transform (PIT) residuals
F) PIT residuals versus ranked model predictions
G) PIT residuals over space

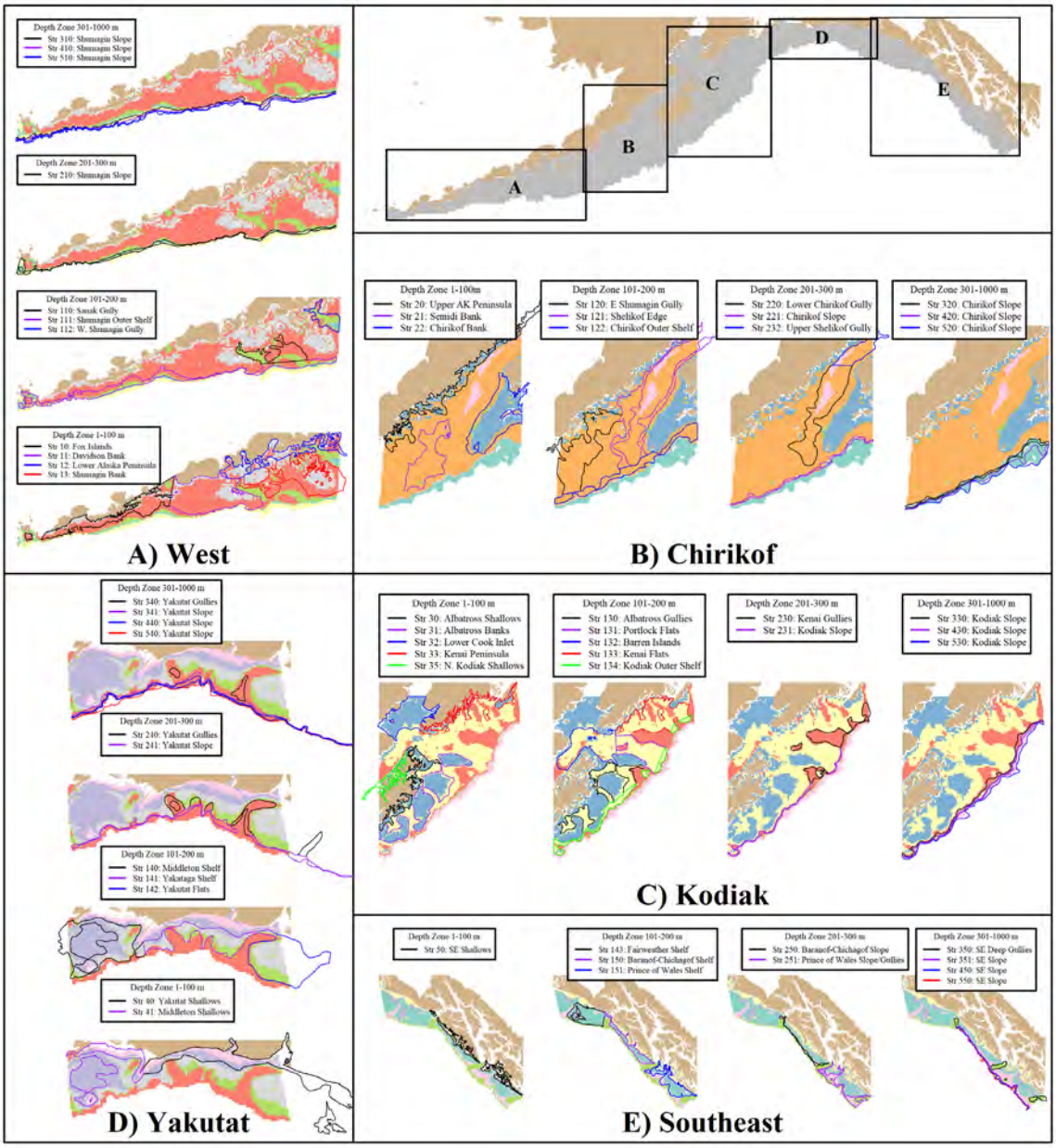
APPENDIX B

Strata boundaries of the existing stratified random survey superimposed on the proposed survey optimization solutions

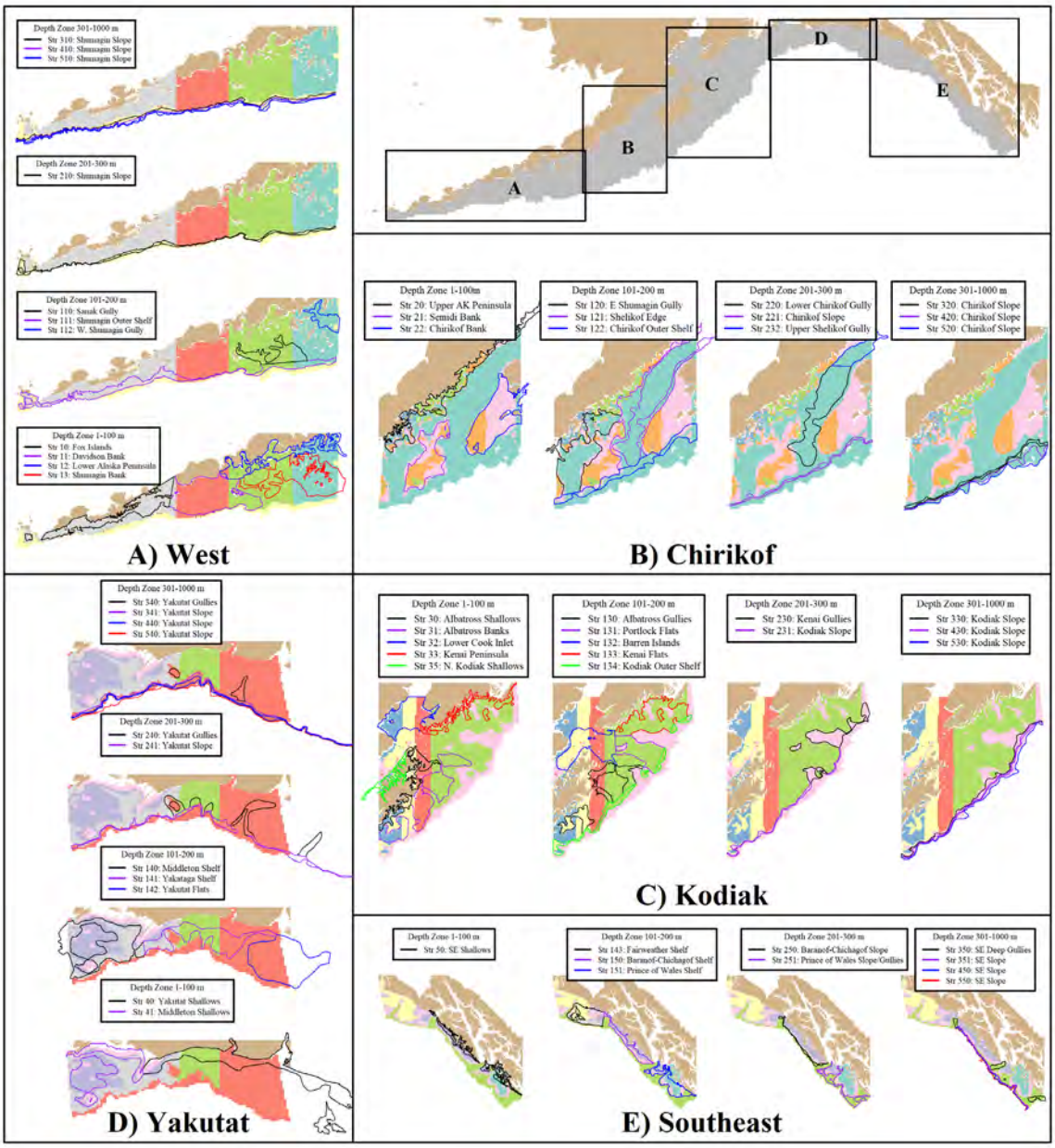
Appendix Figure B-1. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario B (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



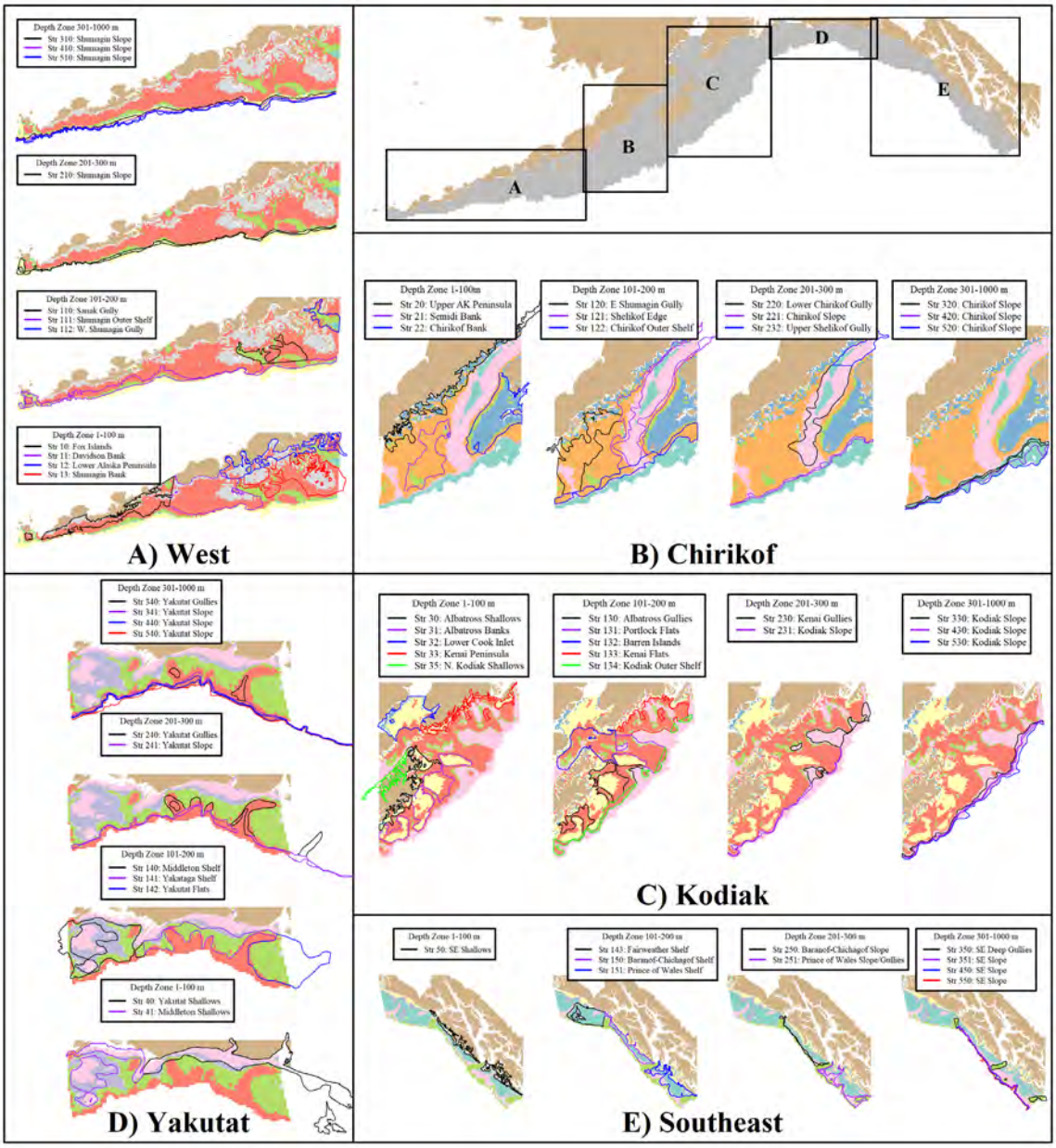
Appendix Figure B-2. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario C (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



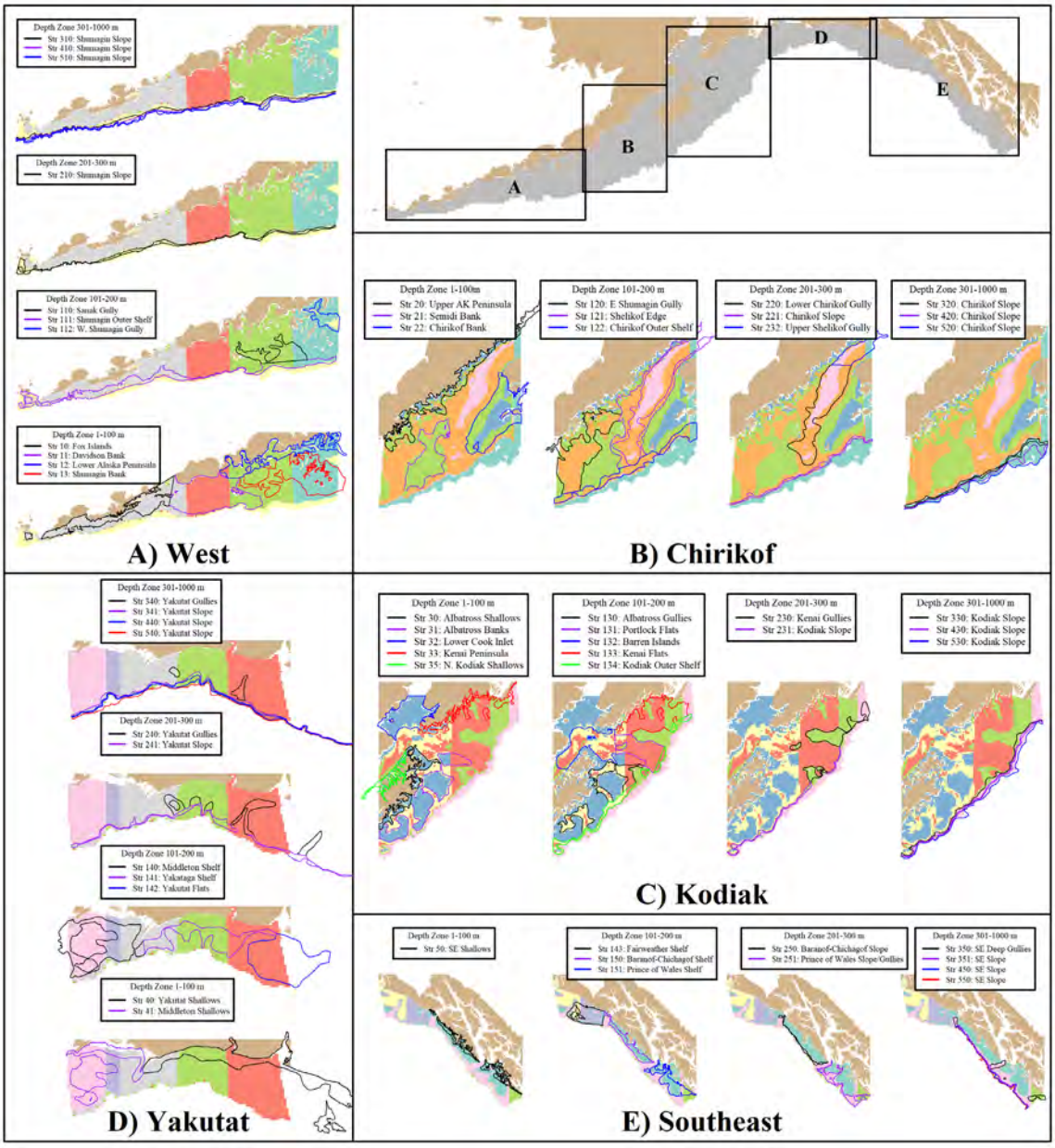
Appendix Figure B-3. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario D (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



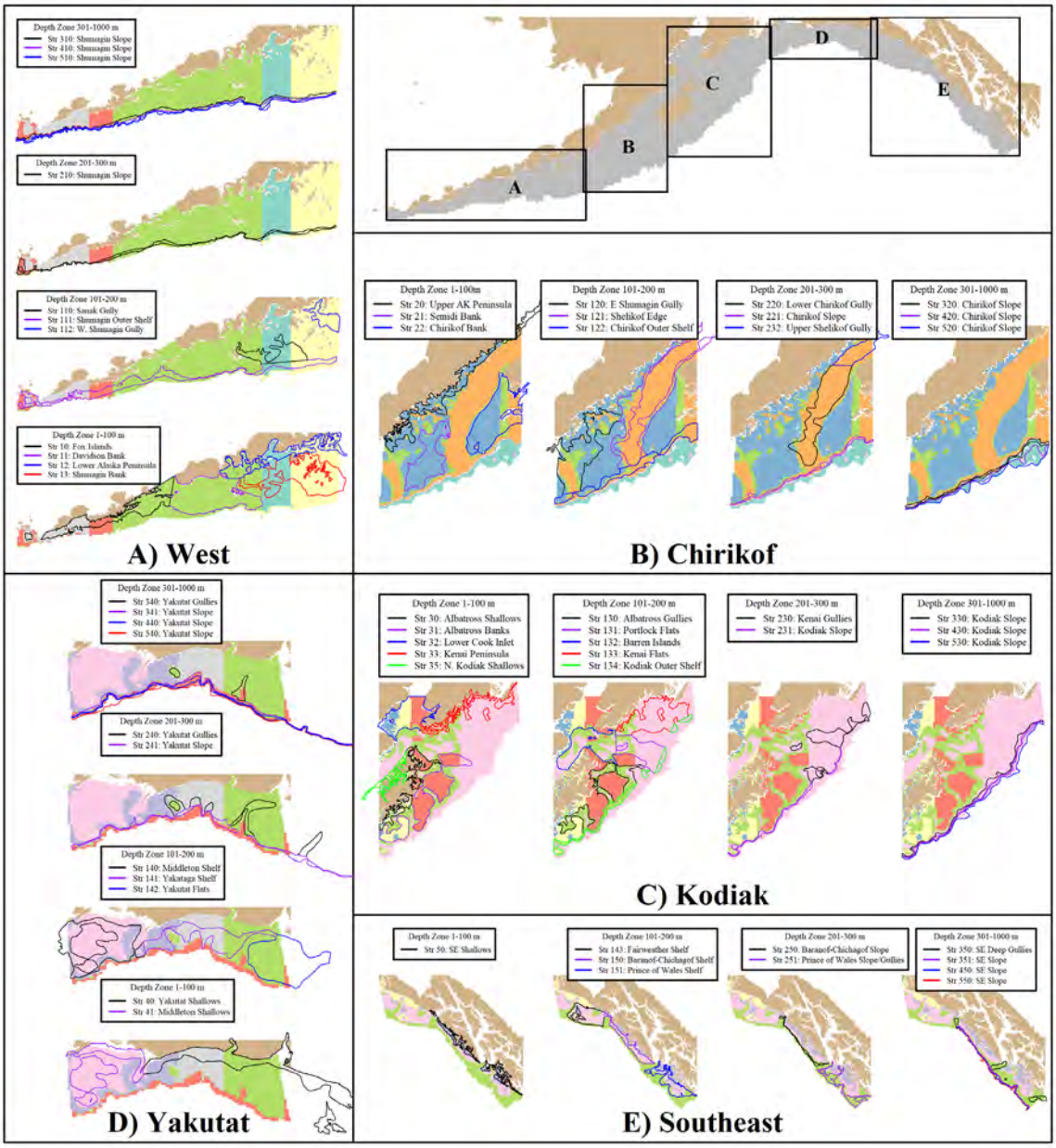
Appendix Figure B-4. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario E (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



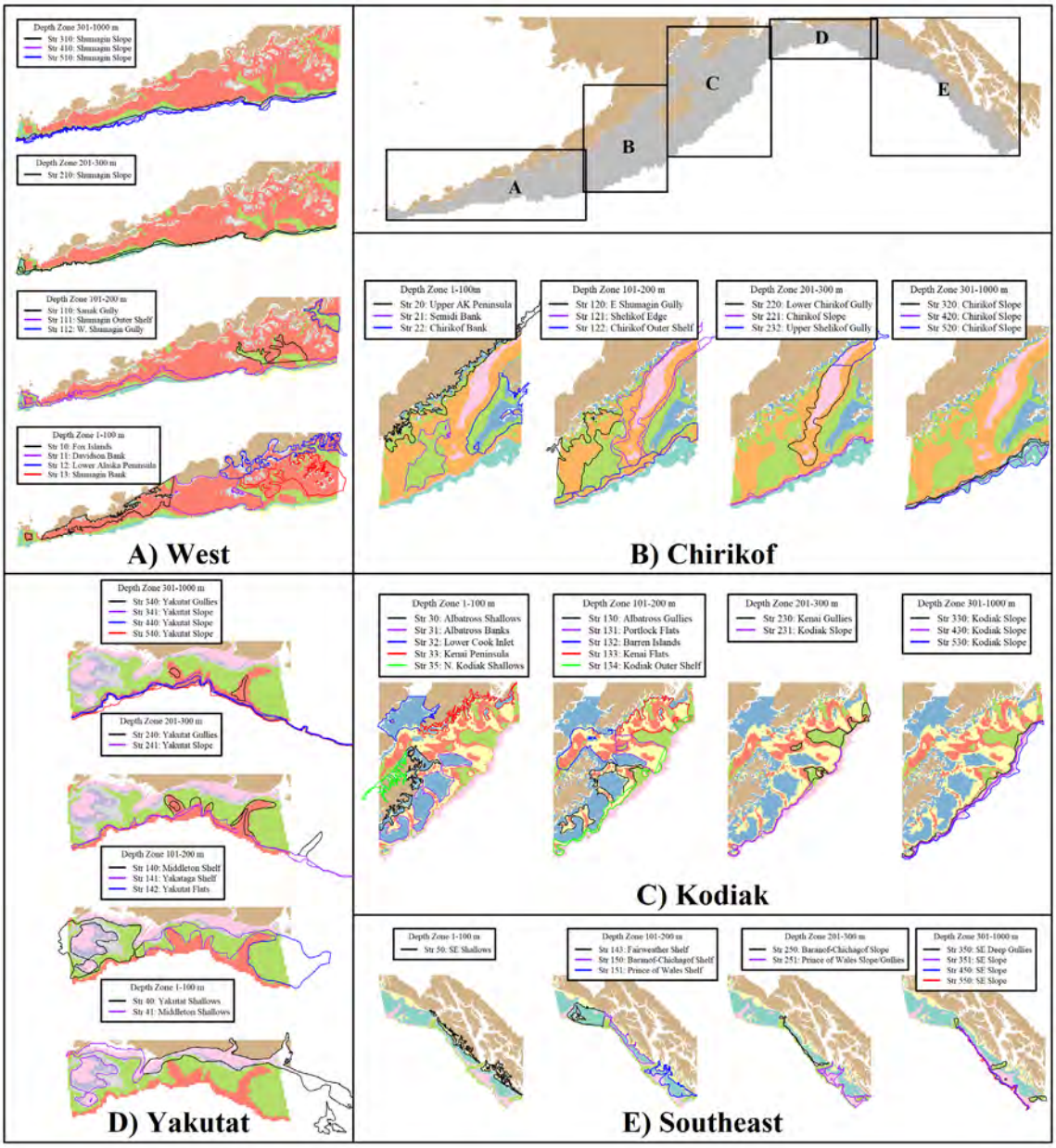
Appendix Figure B-5. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario F (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



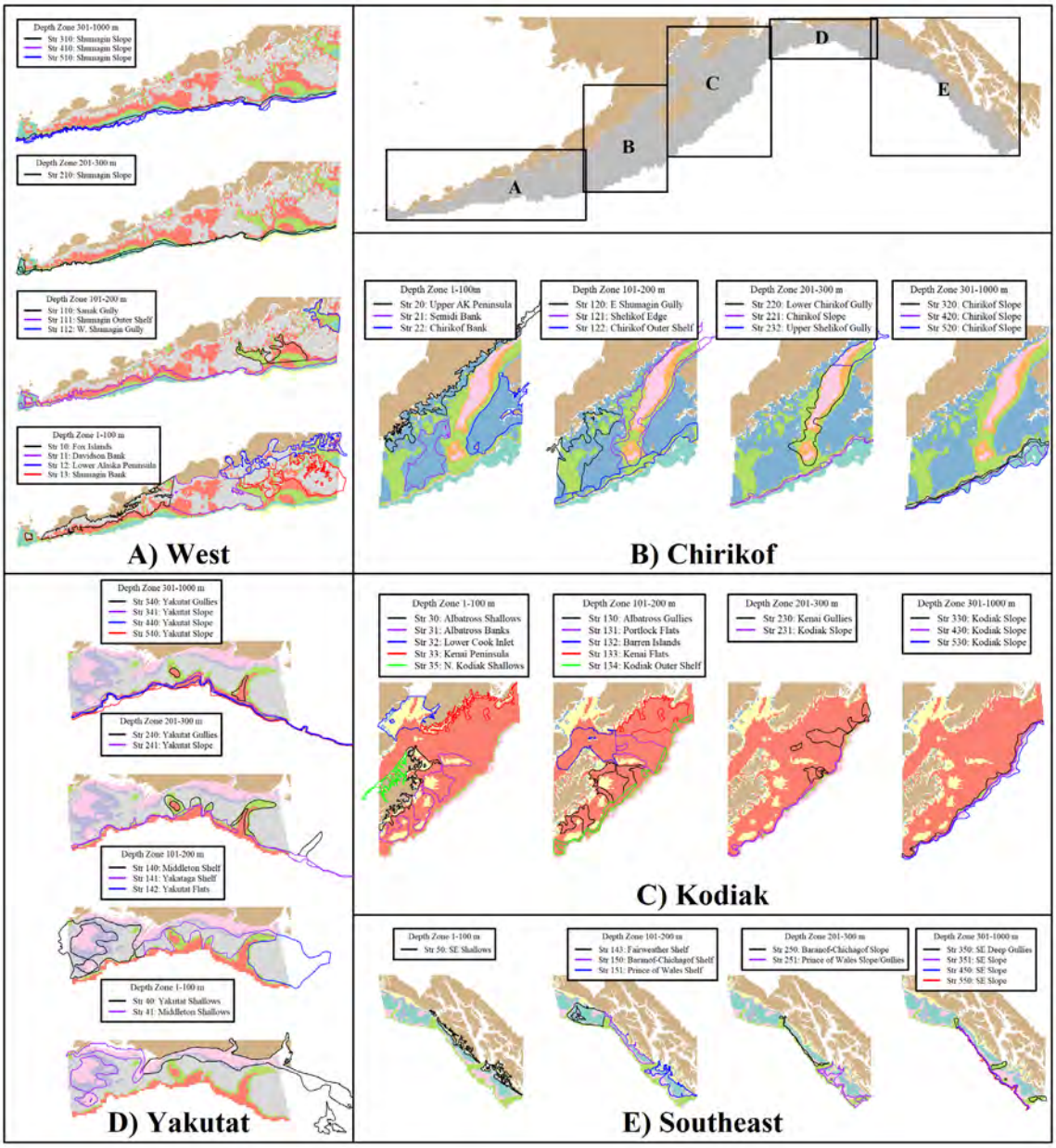
Appendix Figure B-6. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario G (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



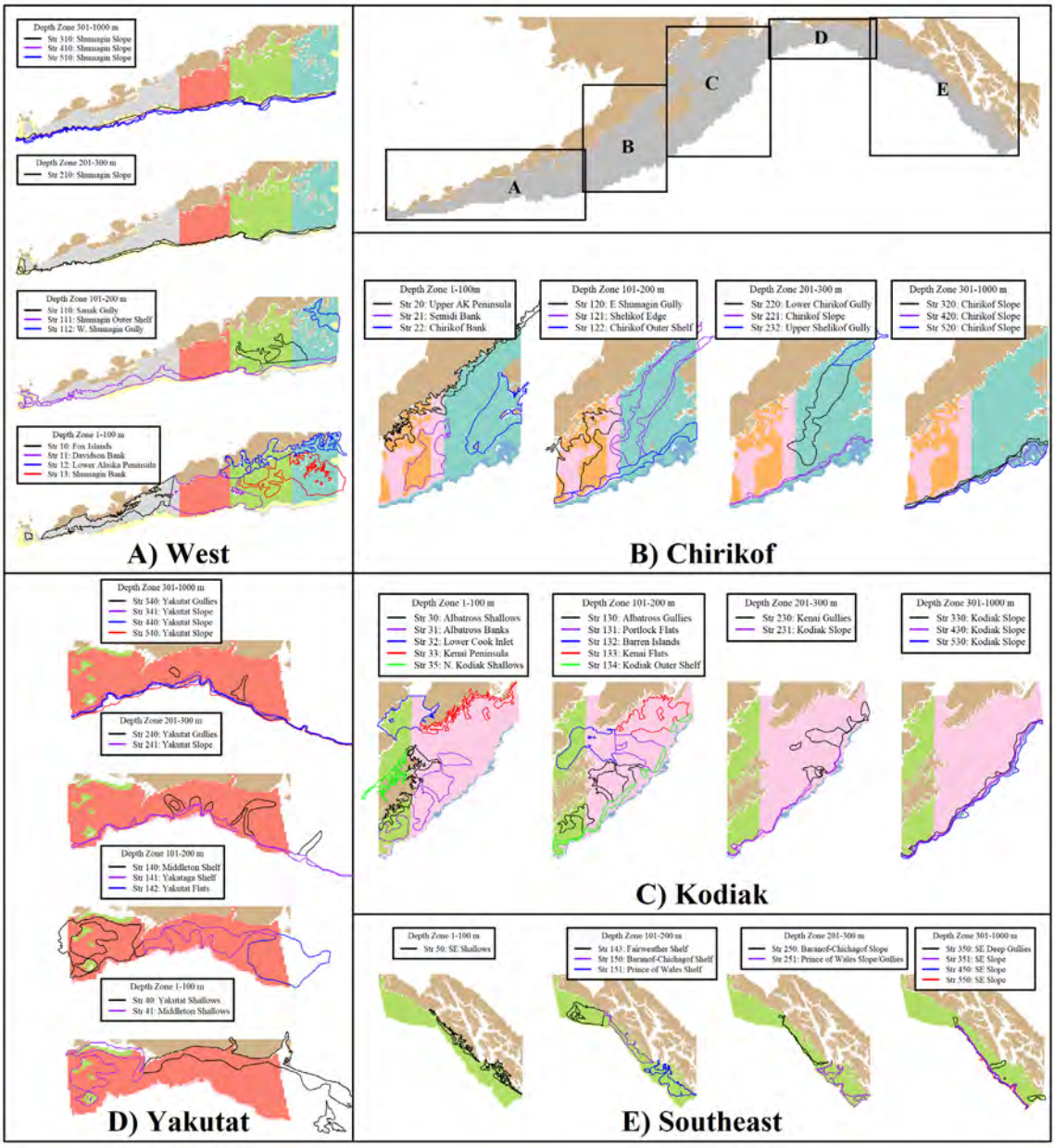
Appendix Figure B-7. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario H (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



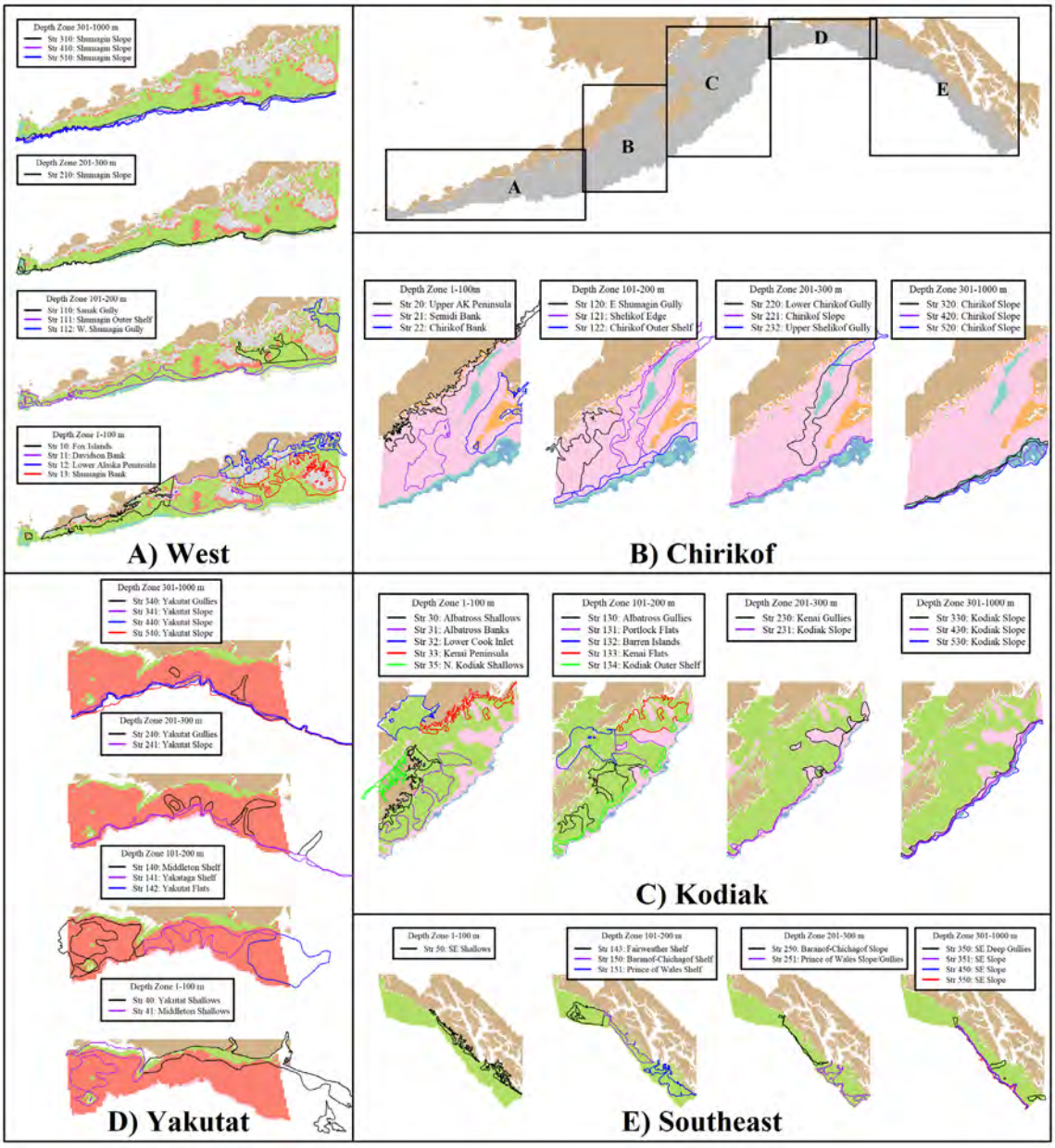
Appendix Figure B-8. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario I (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



Appendix Figure B-9. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario J (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.



Appendix Figure B-10. -- Stratum boundaries of the existing stratified random design (open polygons) superimposed on the proposed optimized strata boundaries under scenario K (five strata per area, filled polygons) defined in Table 4 of the main text. Plots are sectioned by management area.

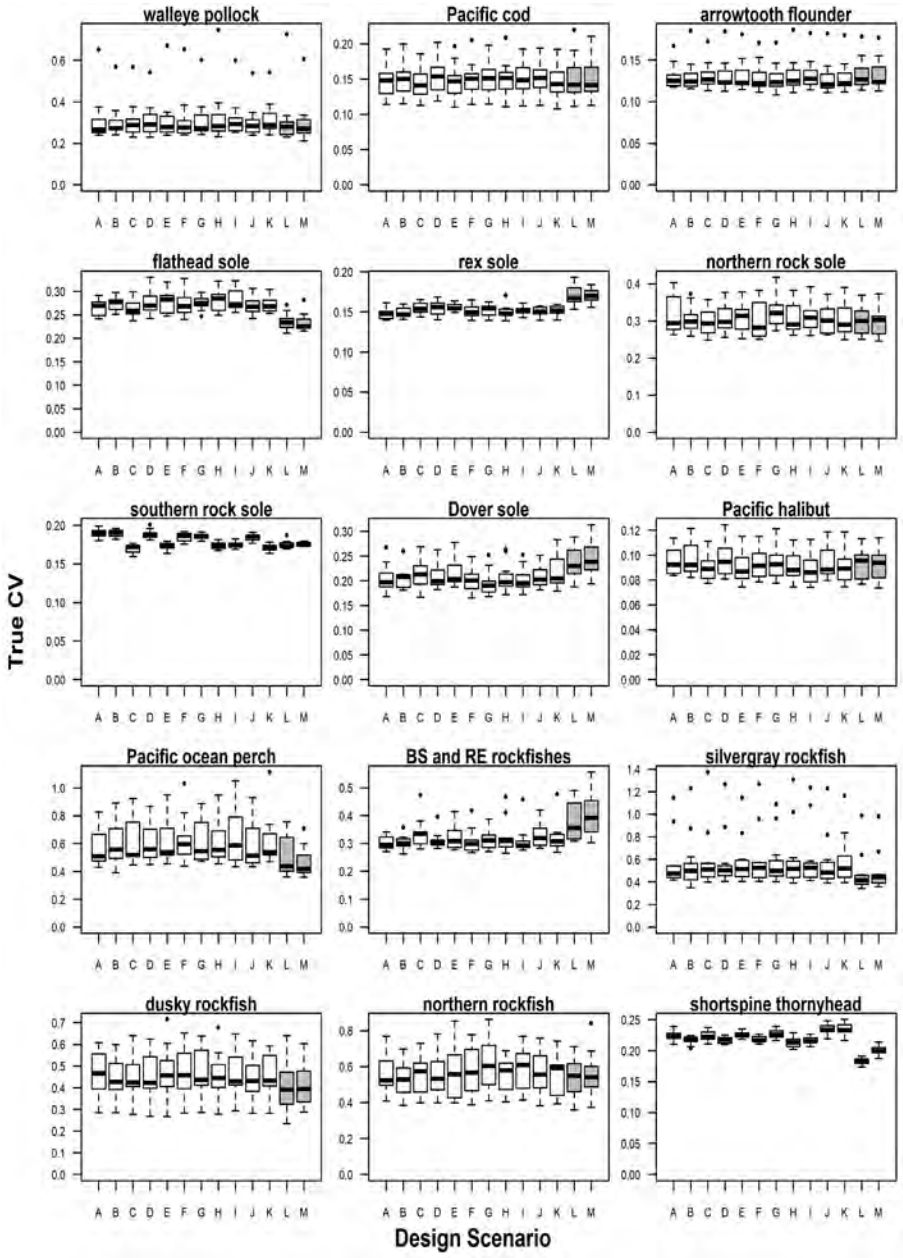


APPENDIX C

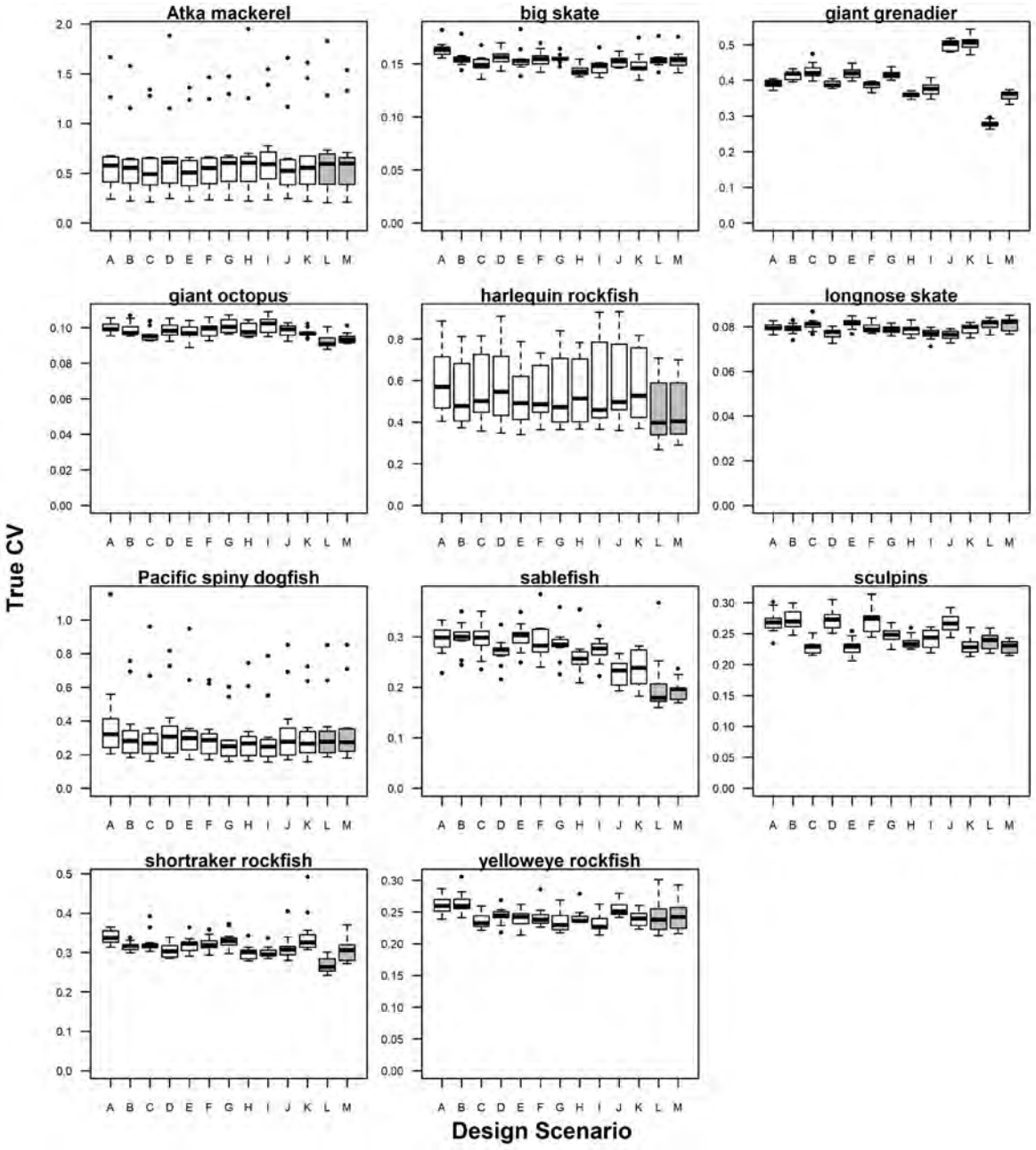
Comparison of Survey Performance Among Designs Given Alternative Total Sample Sizes

Appendix Figures C1-C4 are similar to Figures 8 (True CV) and 9 (relative root mean square error of CV), and Appendix Figures C5 and C6 are similar to Figure 6 (relative bias) in the main text, shown for one and three levels of boat effort. Appendix Figure C7 shows the 95th percentile distribution of the ratio of the abundance index calculated under a proposed design (scenario K as defined in Table 4 of the main text) to the existing survey design (scenario M as defined in Table 4 of the main text) for each species where only cells shallower than 700 m in the spatial domain are included.

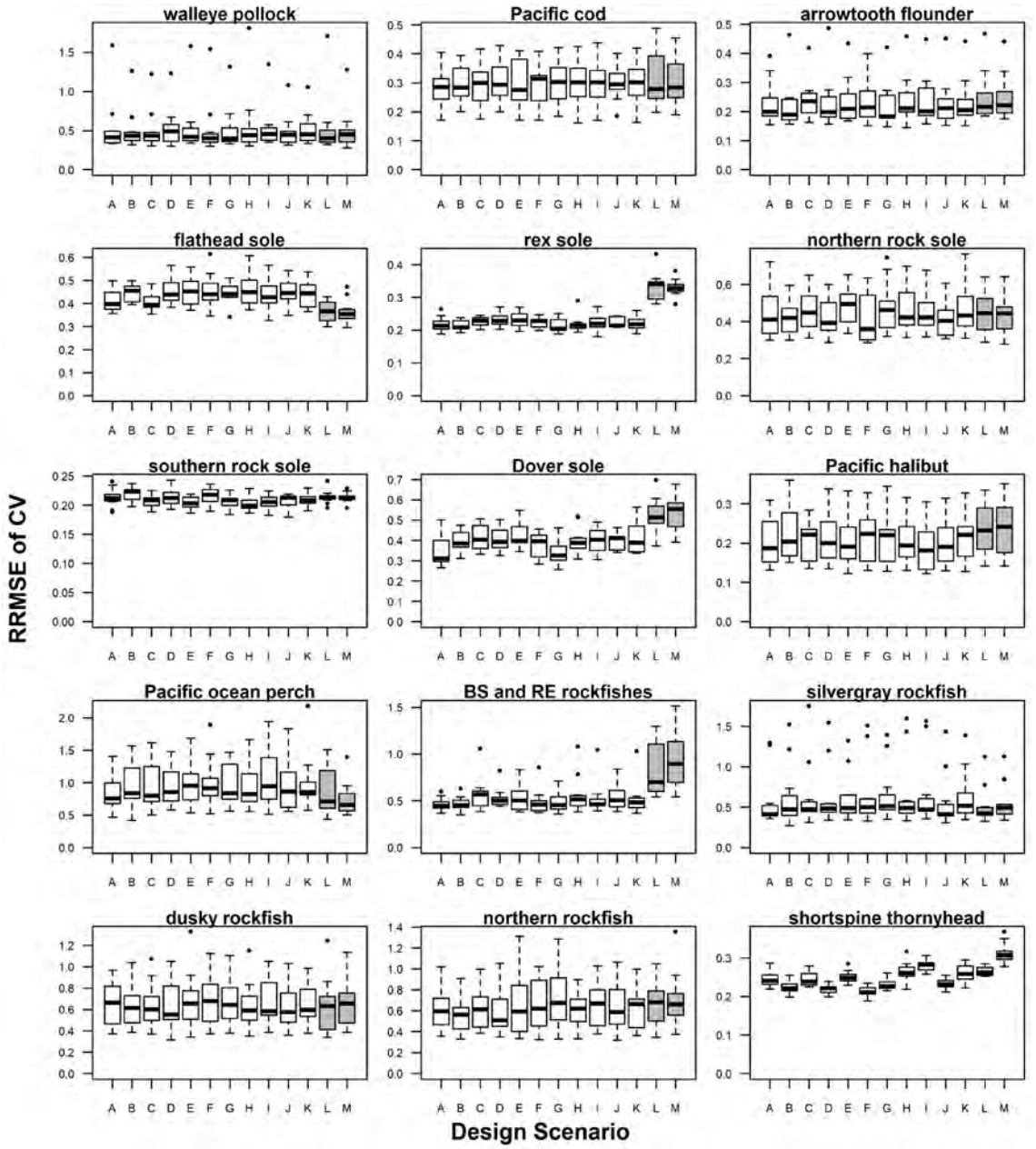
Appendix Figure C-1. -- Boxplot distributions of true coefficient of variation (CV; standard deviation of estimated abundance indices across 1,000 realized survey simulations divided by the true abundance index) across years for each design scenario described in Table 3. Gulf-scale proposed designs (Scenario A) include 15 strata, area-level designs (Scenarios B-K) include five strata per area, and scenarios L-M utilize the existing STRS design. One-boat operations are shown for comparison. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.



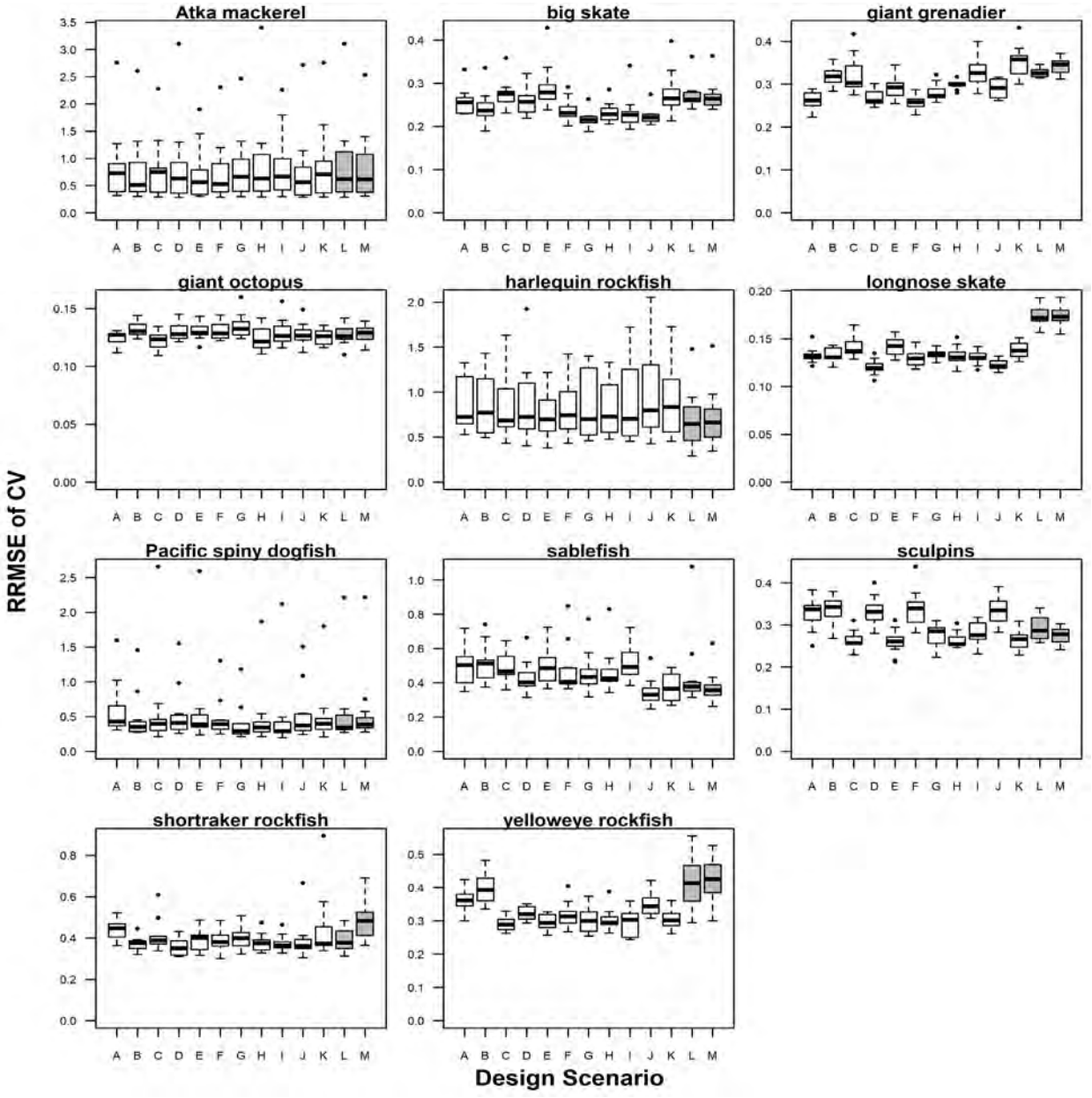
Appendix Figure C-1. -- Continued.



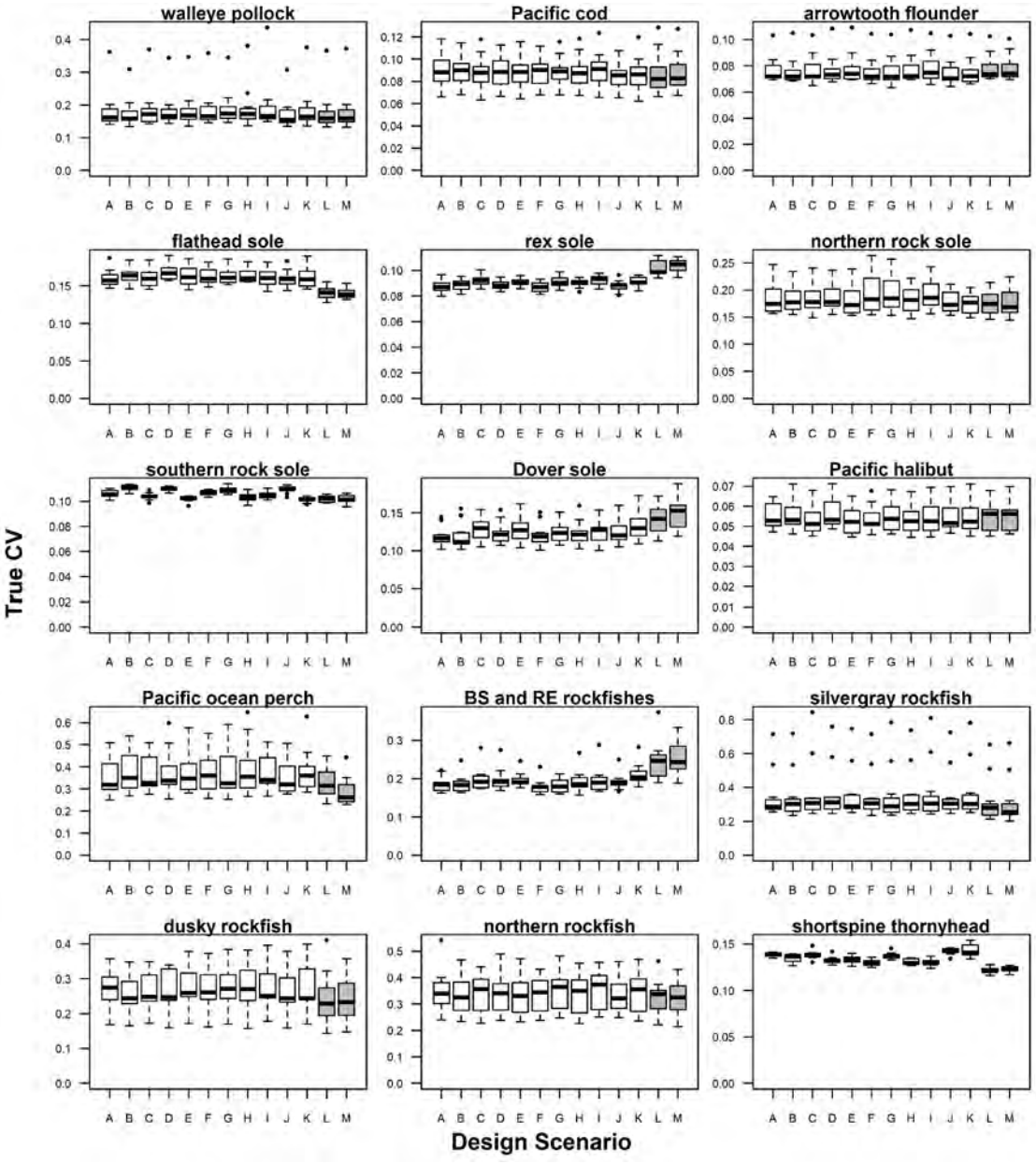
Appendix Figure C-2. -- Boxplot distributions of the relative root mean square error (RRMSE) of the coefficient of variation (CV) across years for each design scenario described in Table 3. Gulf-scale proposed designs (Scenario A) include 15 strata, area-level designs (Scenarios B-K) include five strata per area, and scenarios L-M utilize the existing STRS design. Two-boat operations are shown for comparison. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.



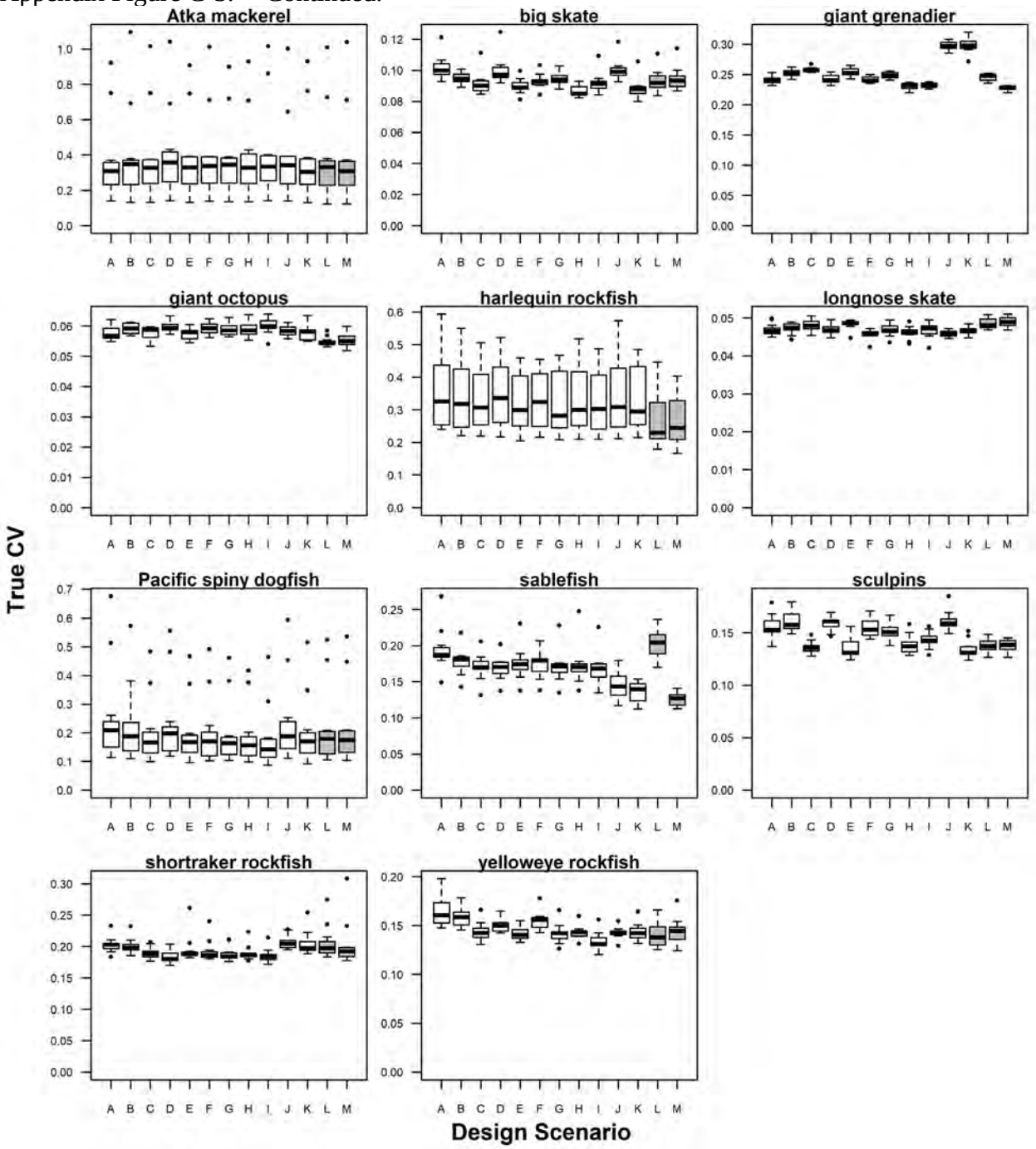
Appendix Figure C-2. -- Continued.



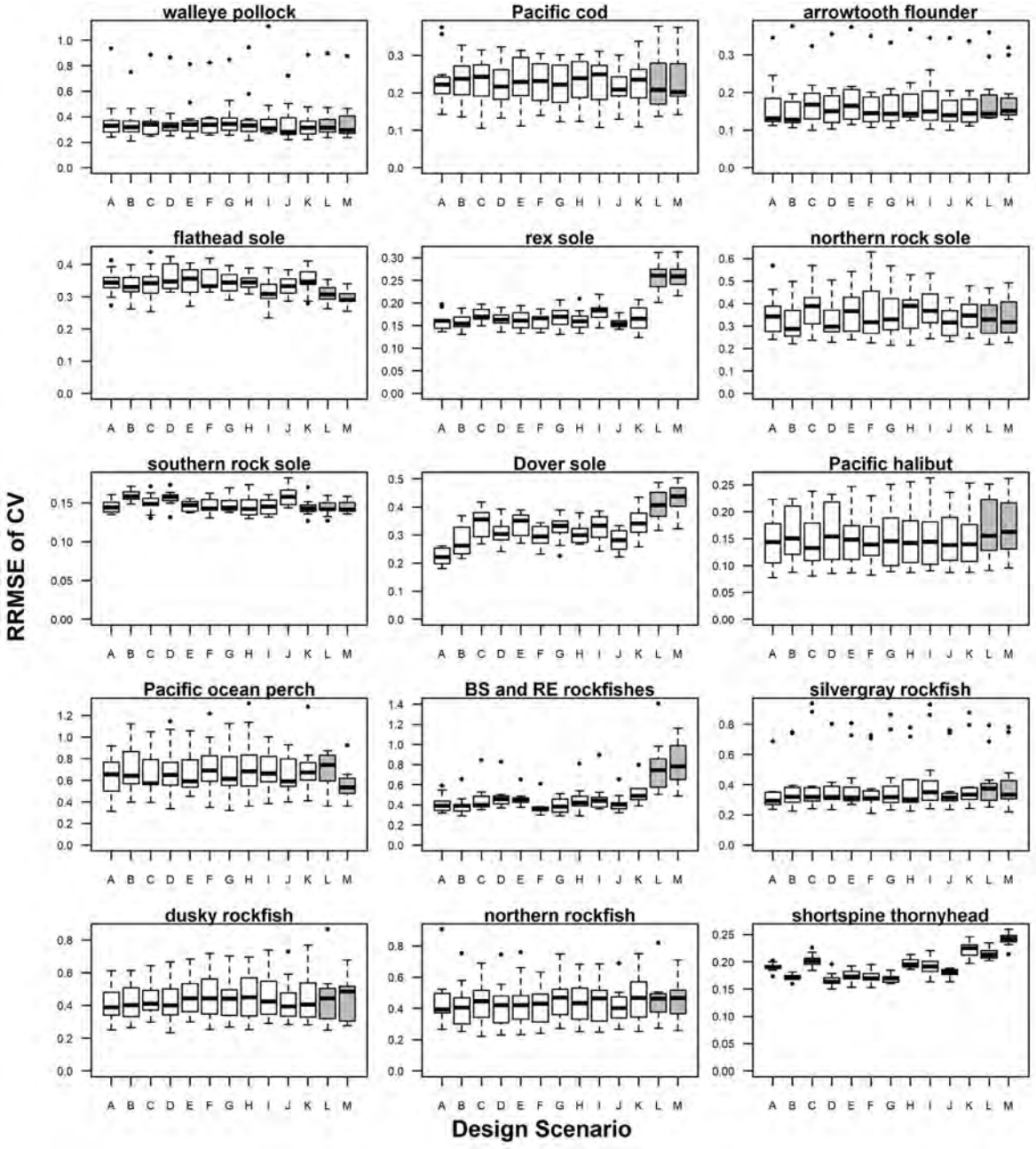
Appendix Figure C-3. -- Boxplot distributions of true coefficient of variation (CV; standard deviation of estimated abundance indices across 1,000 realized survey simulations divided by the true abundance index) across years for each design scenario described in Table 3. Gulf-scale proposed designs (Scenario A) include 15 strata, area-level designs (Scenarios B-K) include five strata per area, and scenarios L-M utilize the existing STRS design. Three-boat operations are shown for comparison. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.



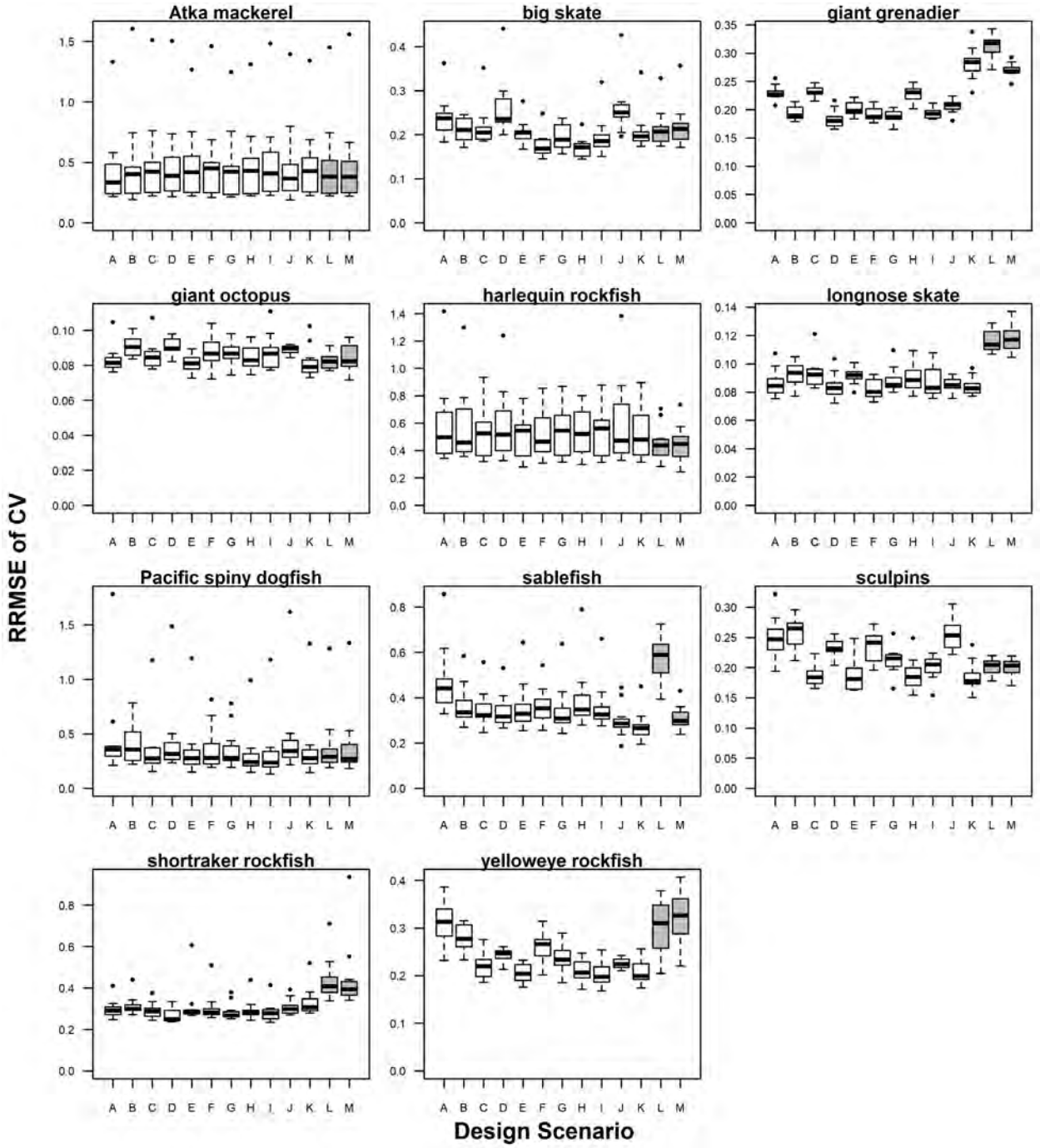
Appendix Figure C-3. -- Continued.



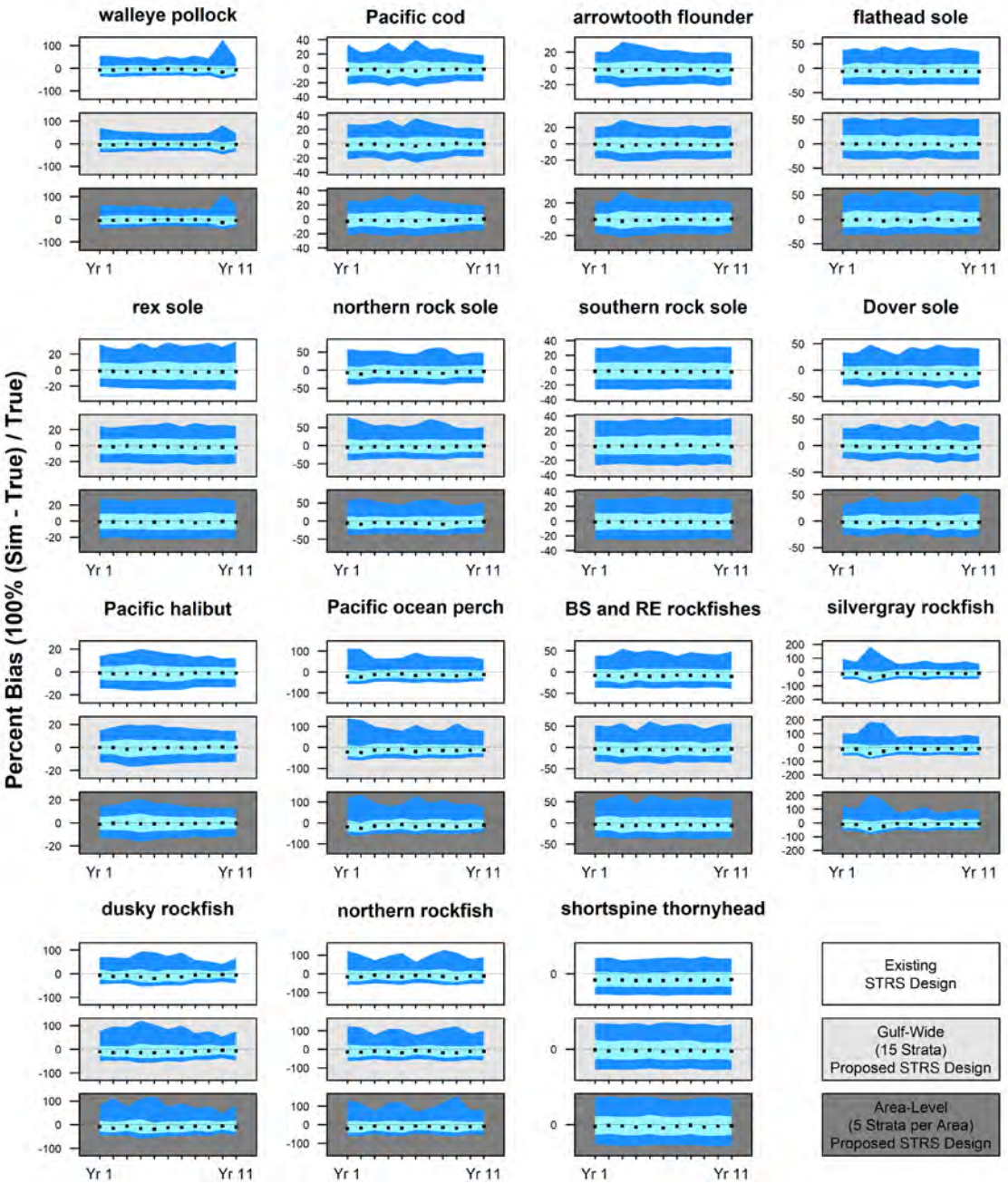
Appendix Figure C-4. -- Boxplot distributions of the relative root mean square error (RRMSE) of the coefficient of variation (CV) across years for each design scenario described in Table 3. Gulf-scale proposed designs (Scenario A) include 15 strata, area-level designs (Scenarios B-K) include five strata per area, and scenarios L-M utilize the existing STRS design. Two-boat operations are shown for comparison. Whisker length indicates 1.5 times the interquartile range, lower and upper hinges correspond to the 25th and 75th percentiles.



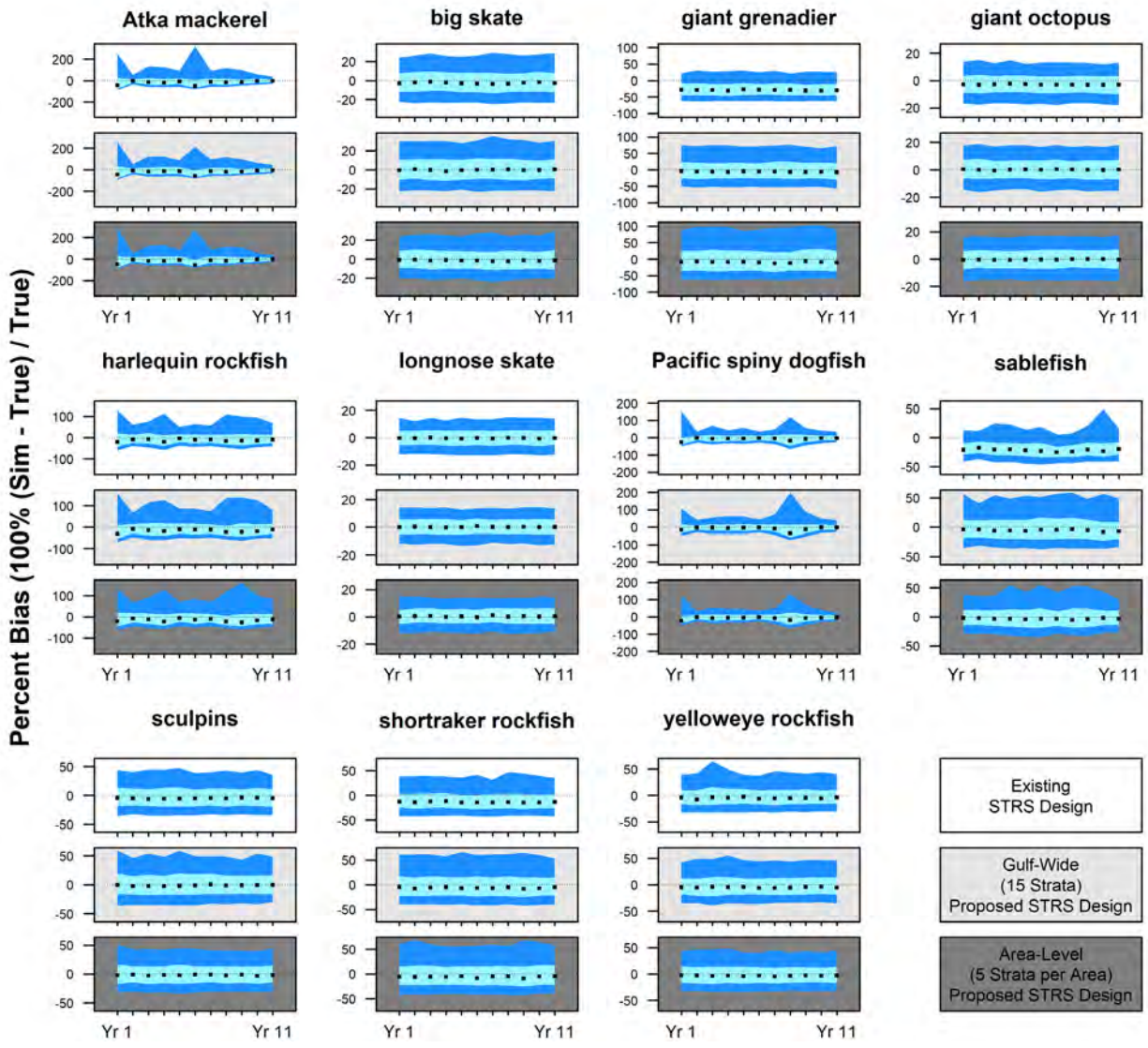
Appendix Figure C-4. -- Continued.



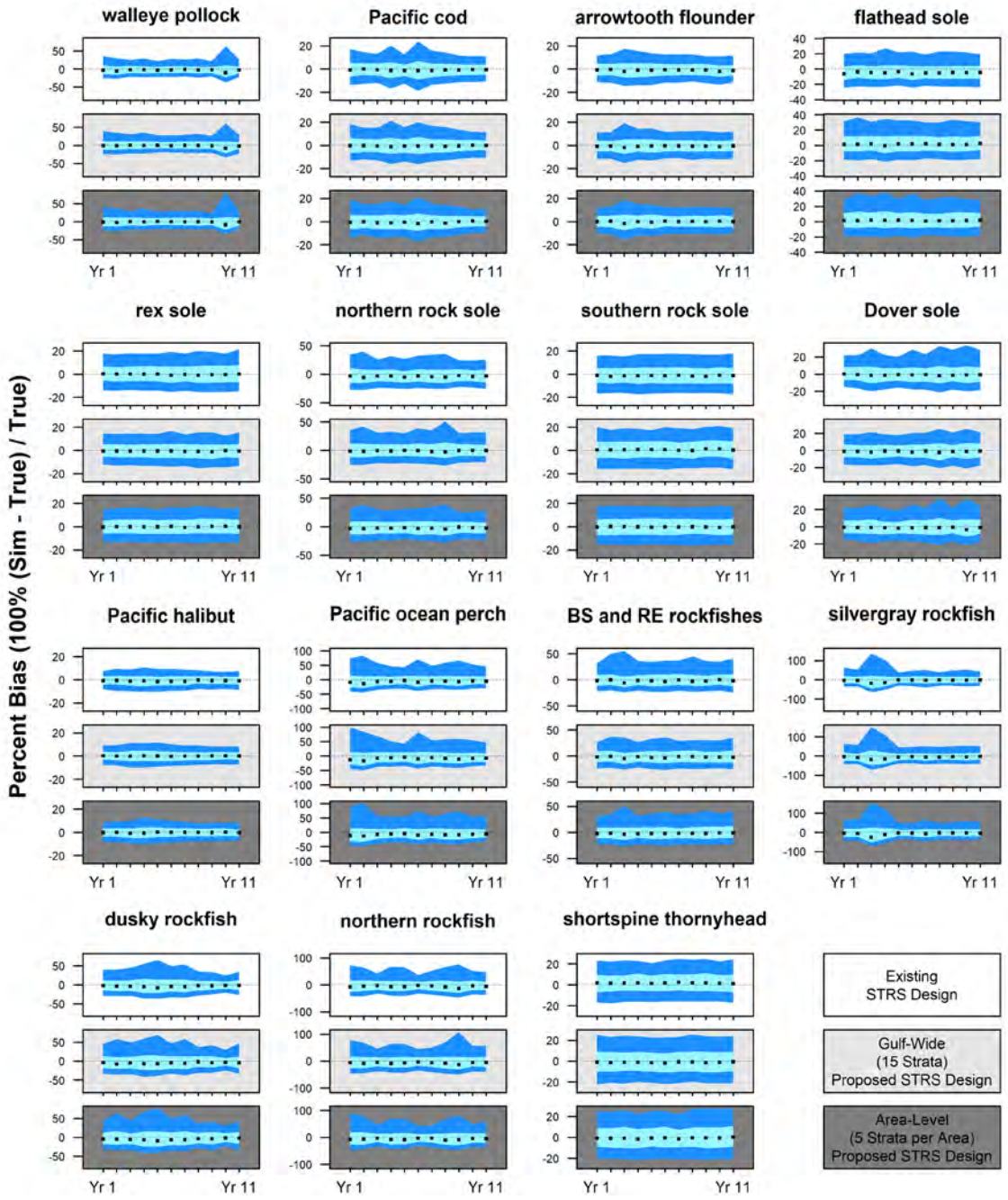
Appendix Figure C-5. -- Median (with 90% percentile confidence interval) relative bias of simulated total (gulf-wide) abundance indices across time (x-axis) and species. Color rows indicate different surveys: existing survey (white), gulf-wide optimized survey under scenario A defined in Table 4 (light grey, 15 strata), and area-level optimized survey under scenario B defined in Table 4 (dark grey, five strata per area). All survey scenarios were one-boat effort operations for comparison.



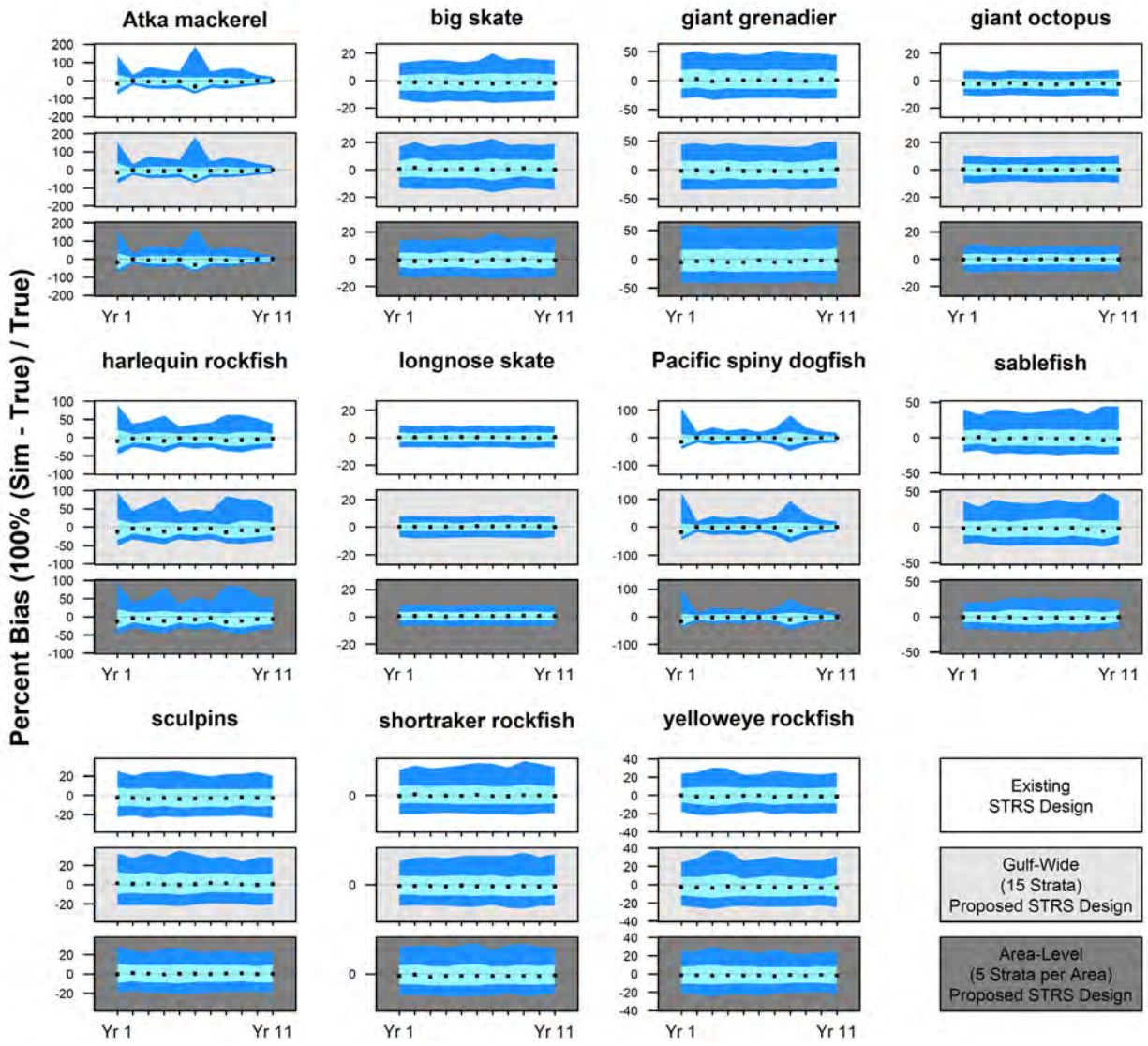
Appendix Figure C-5. -- Continued.



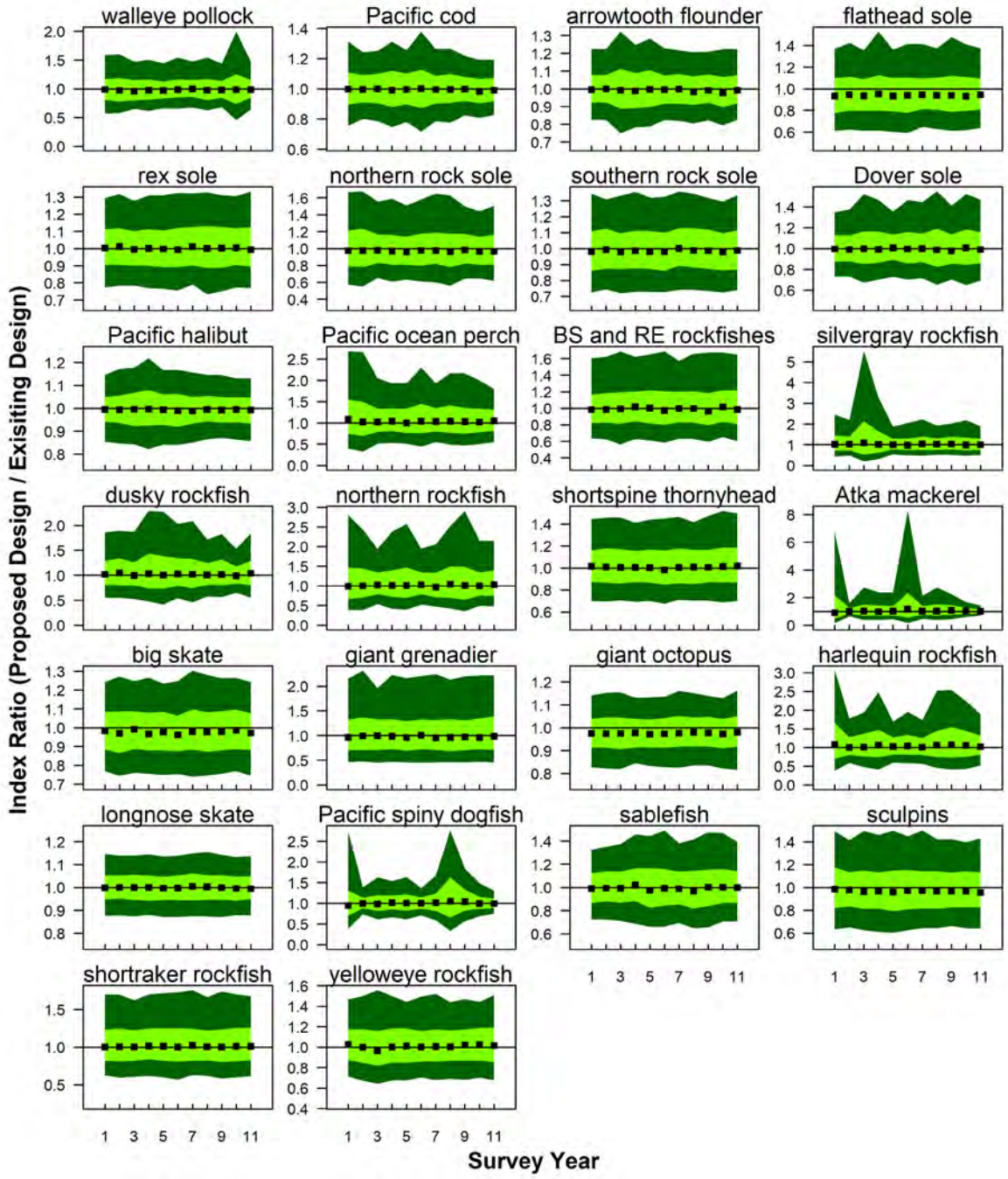
Appendix Figure C-6. -- Median (with 90% percentile confidence interval) relative bias of simulated total (gulf-wide) abundance indices across time (x-axis) and species. Color rows indicate different surveys: existing survey (white), gulf-wide optimized survey under scenario A defined in Table 4 (light grey, 15 strata), and area-level optimized survey under scenario B defined in Table 4 (dark grey, five strata per area). All survey scenarios were three-boat effort operations for comparison.



Appendix Figure C-6. -- Continued.



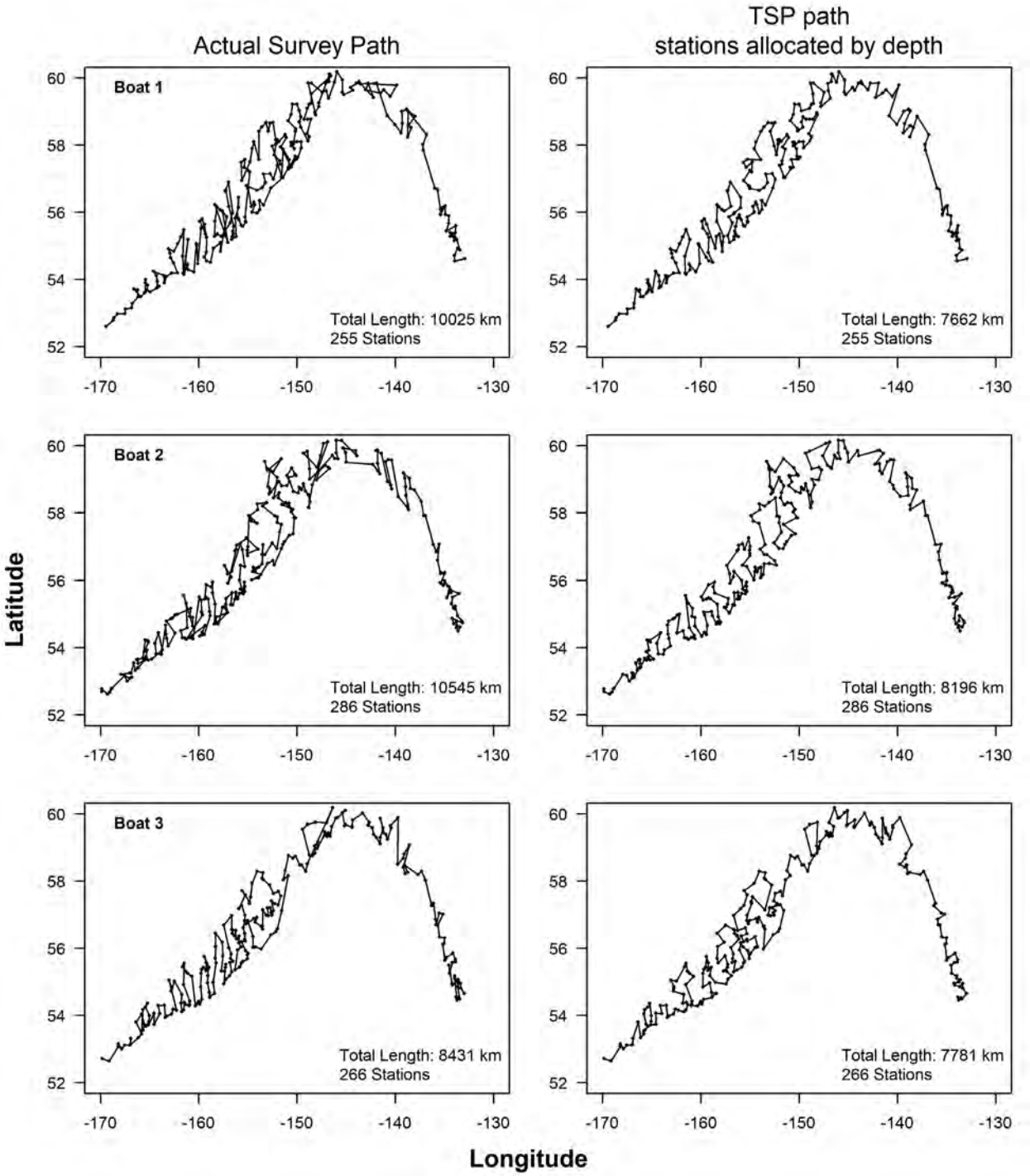
Appendix Figure C-7. -- Median (with 90% percentile confidence interval) ratios of indices produced using 1000 survey replicates of proposed to existing survey designs across survey years (x-axis) and species. Lighter green areas indicate the interquartile region and the darker green encompasses the 90% percentile confidence interval. All survey scenarios were two-boat operations for comparison.



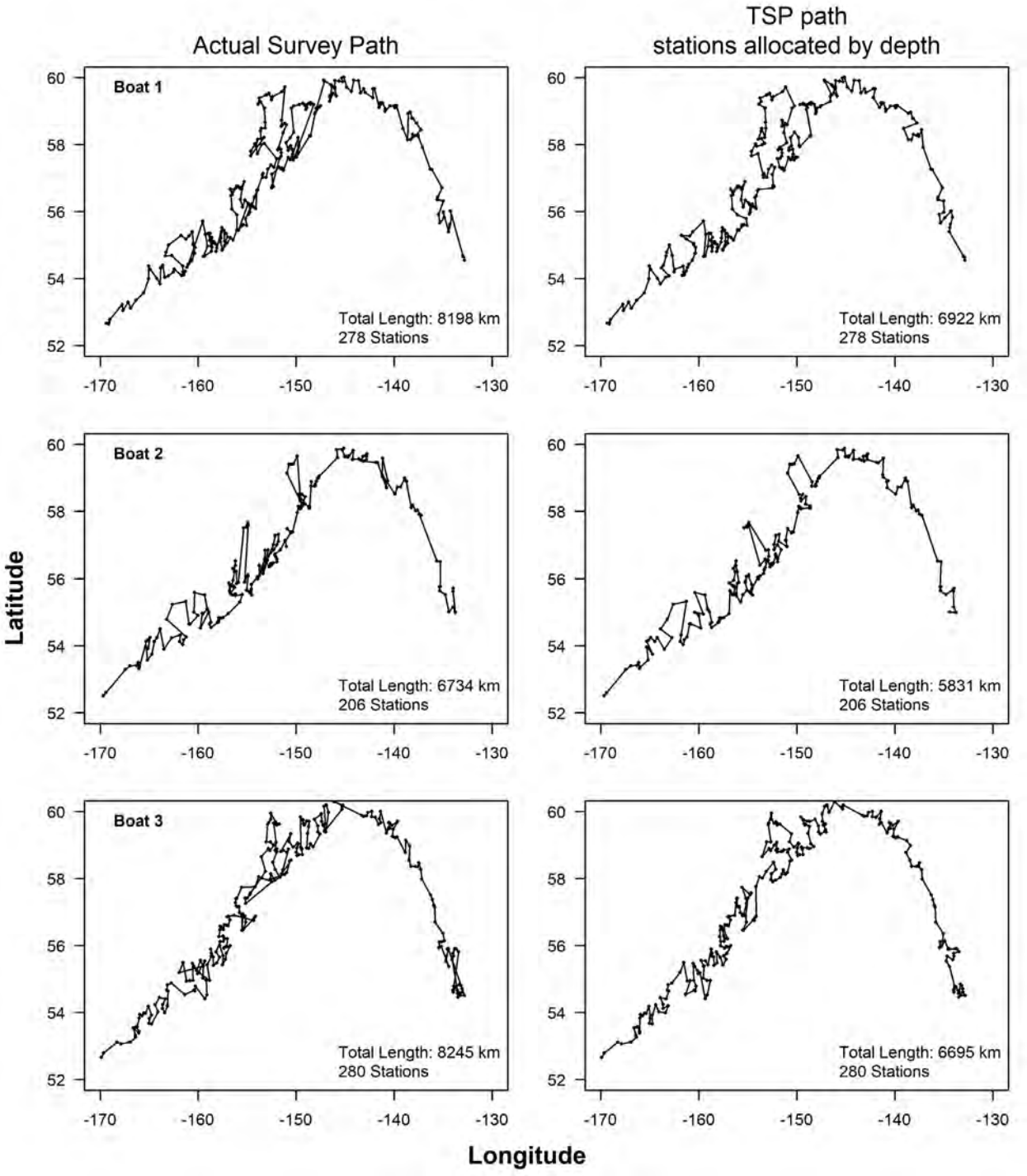
APPENDIX D

Observed Survey Paths and Shortest Paths Calculated by Solving the Travelling Salesperson Problem for Each Boat for a Given Survey Year

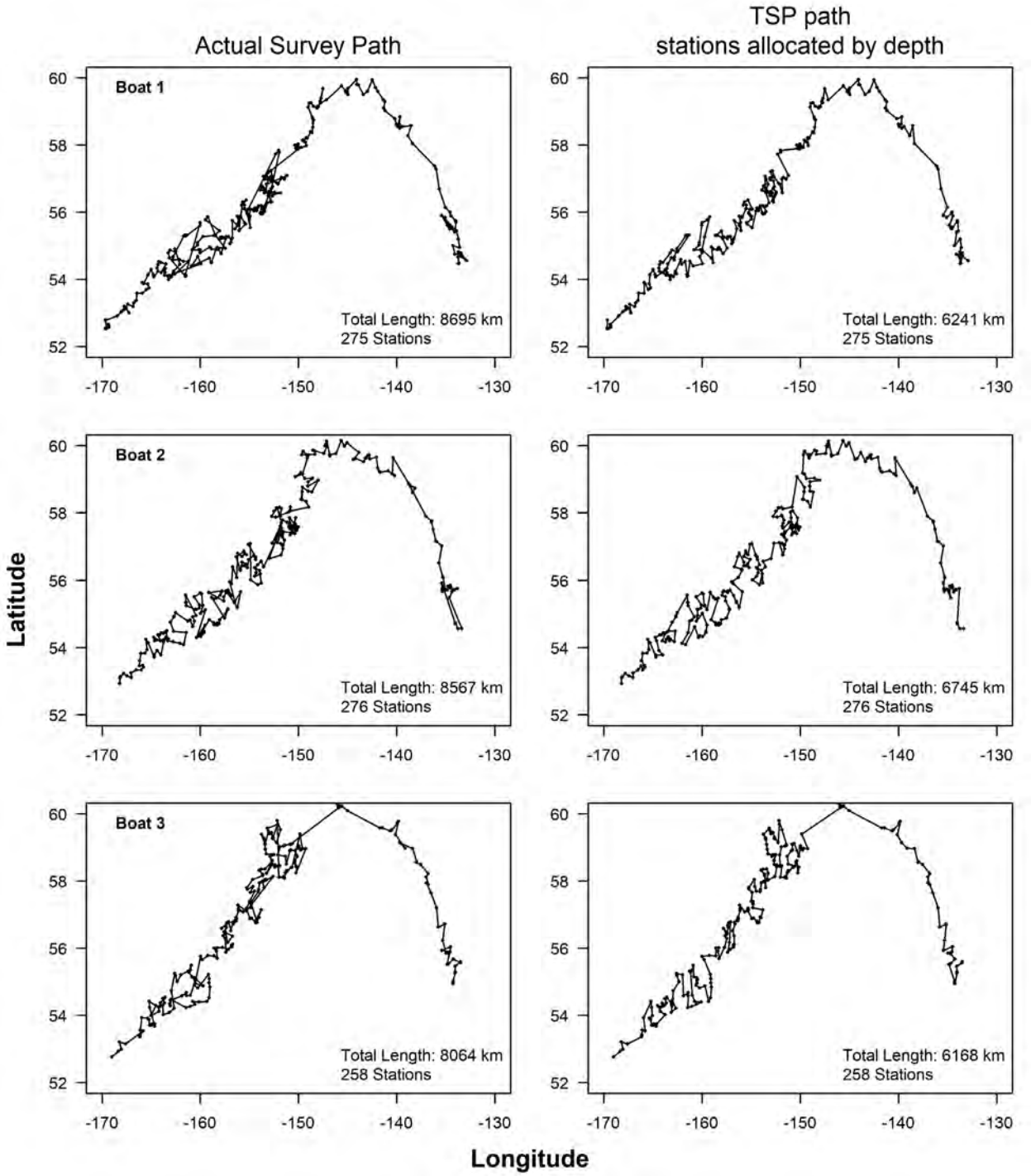
Appendix Figure D-1. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 1996. Surveys start at the most western station and traverse eastward.



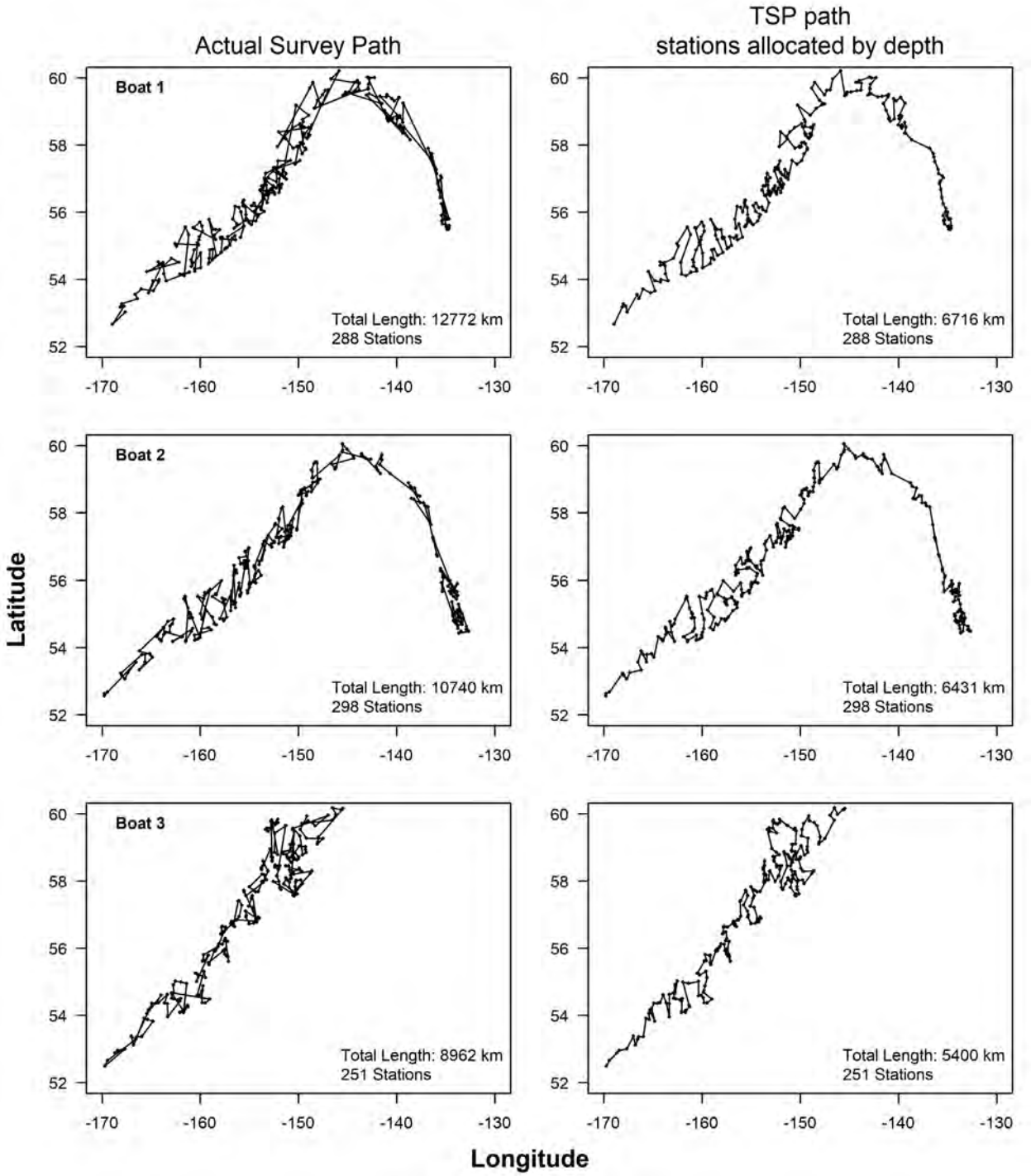
Appendix Figure D-2. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 1999. Surveys start at the most western station and traverse eastward.



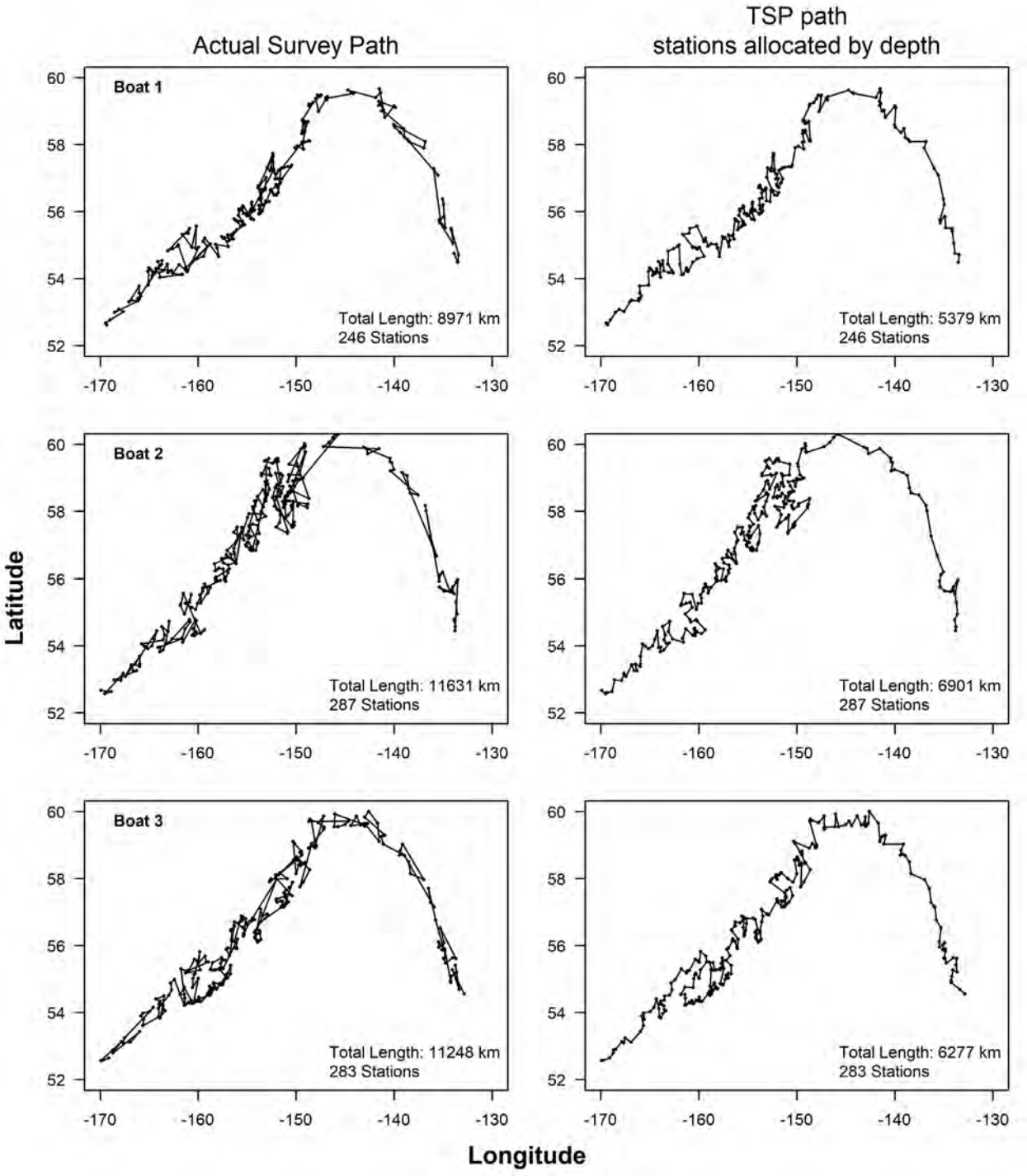
Appendix Figure D-3. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2003. Surveys start at the most western station and traverse eastward.



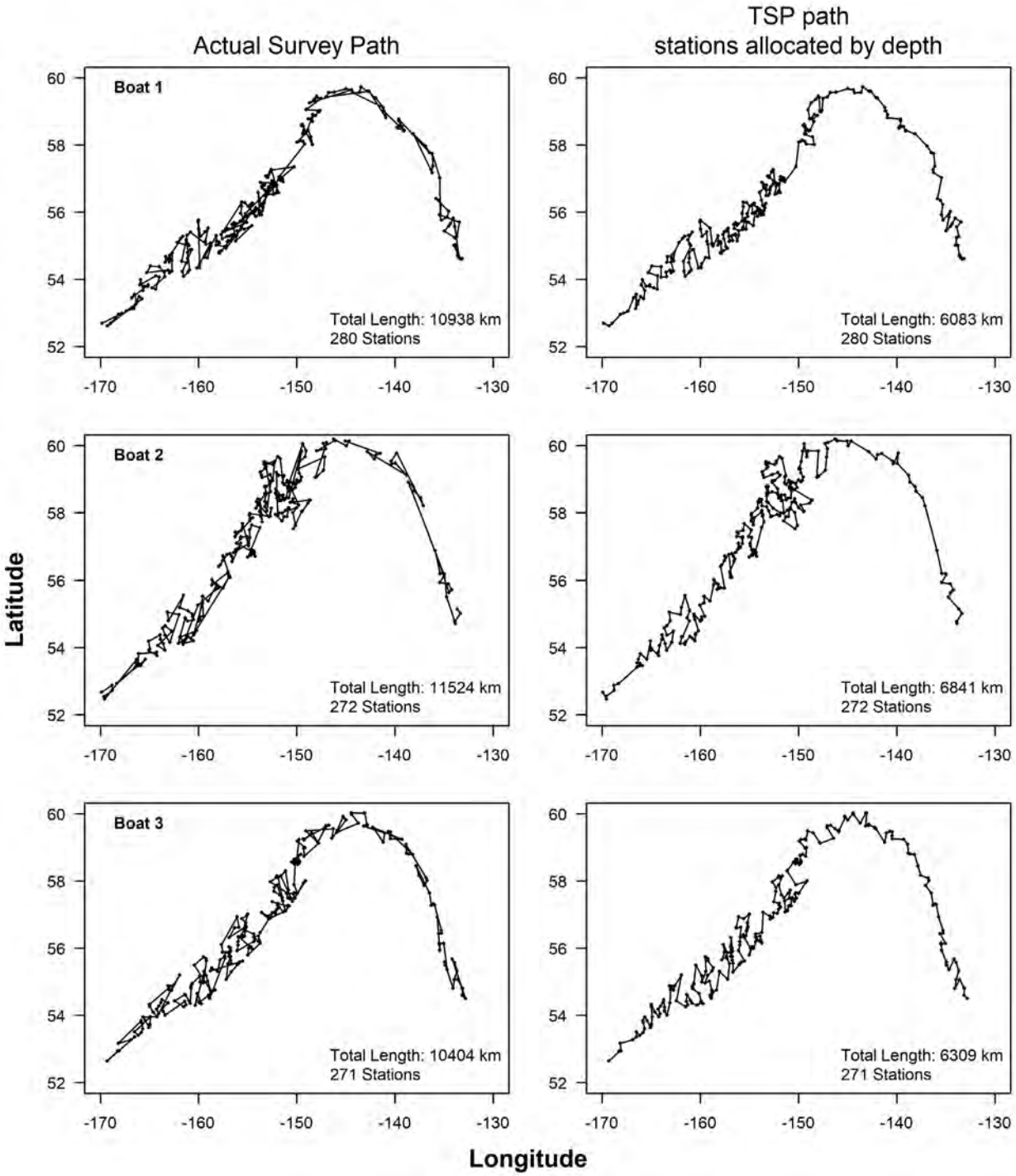
Appendix Figure D-4. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2005. Surveys start at the most western station and traverse eastward.



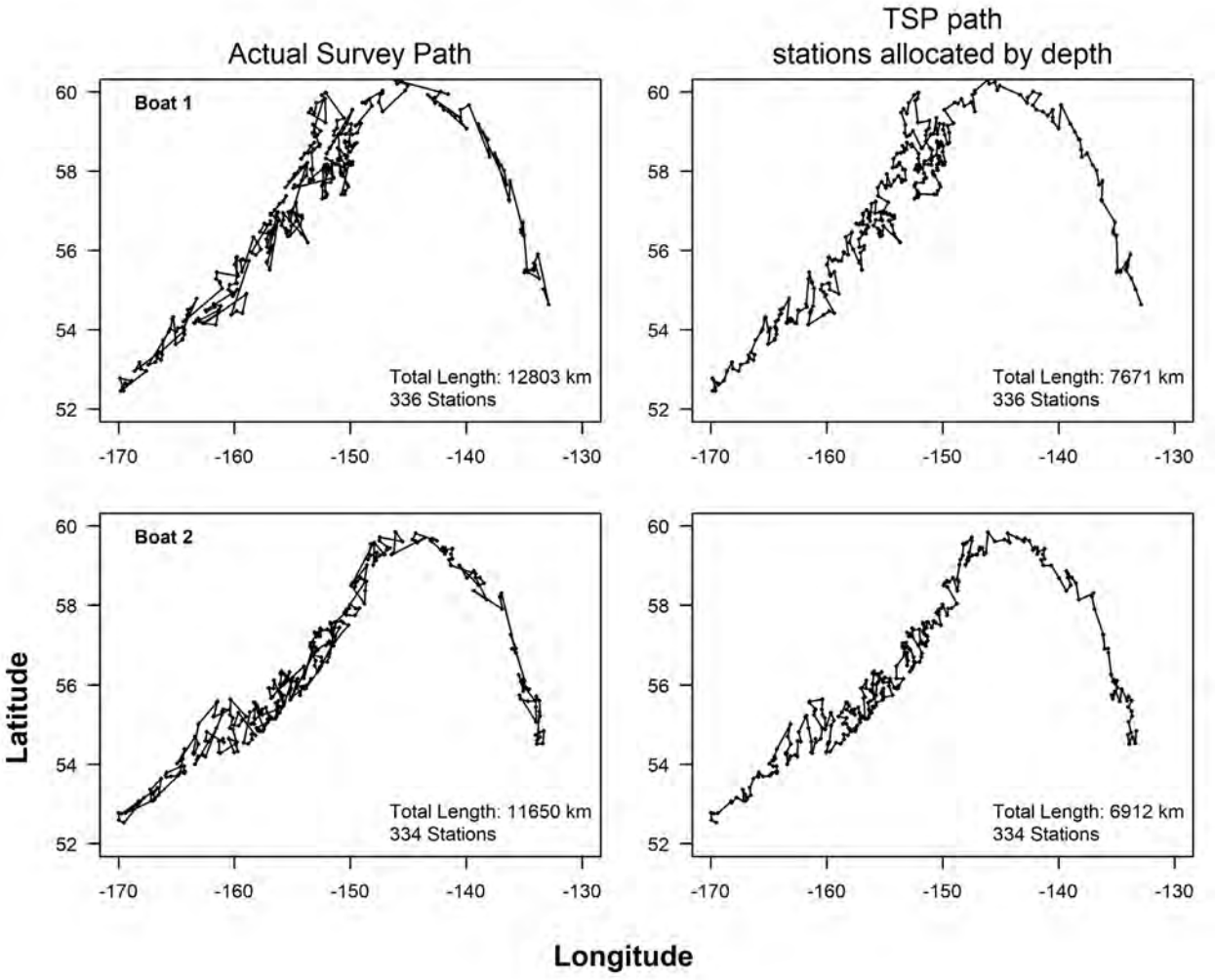
Appendix Figure D-5. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2007. Surveys start at the most western station and traverse eastward.



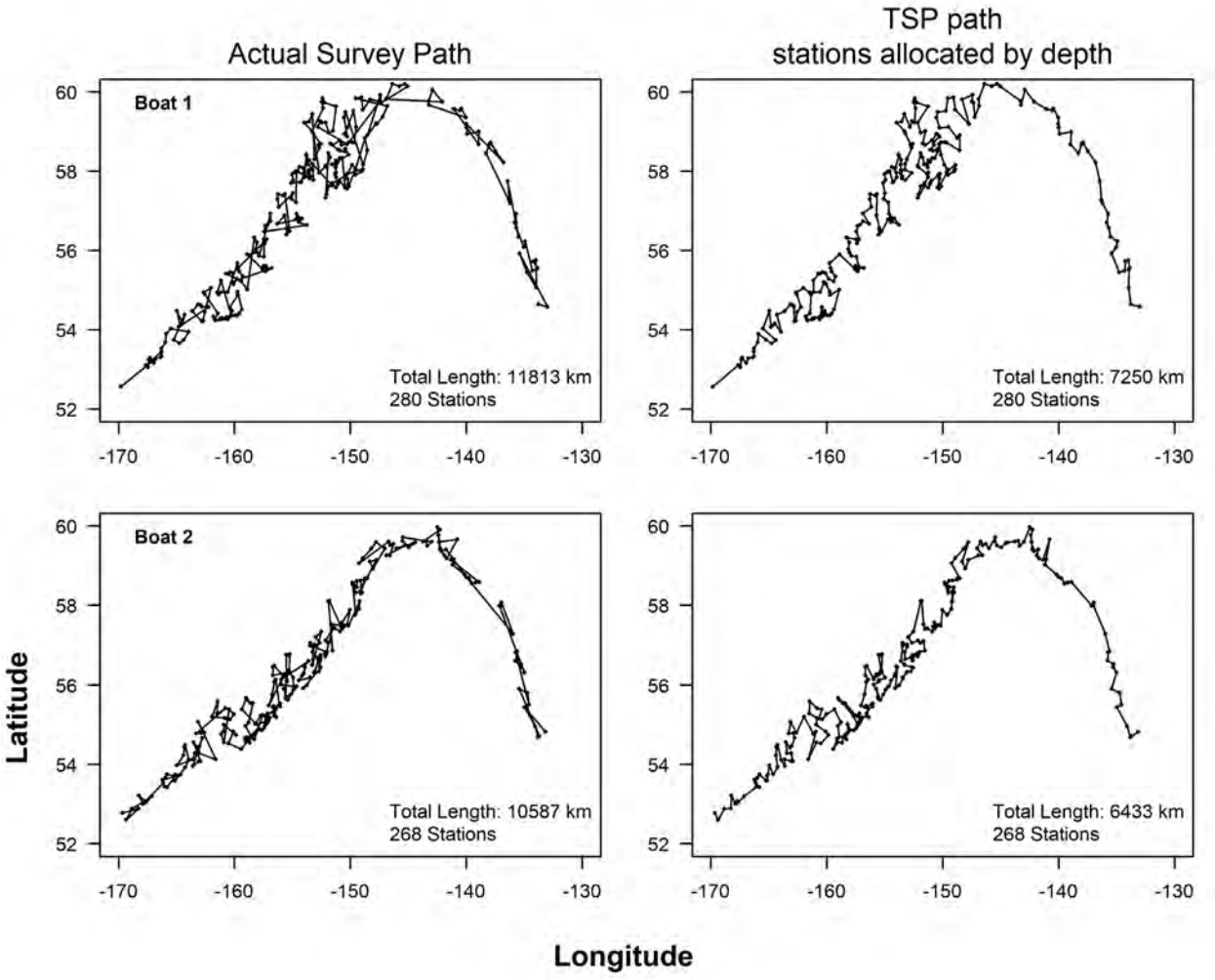
Appendix Figure D-6. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2009. Surveys start at the most western station and traverse eastward.



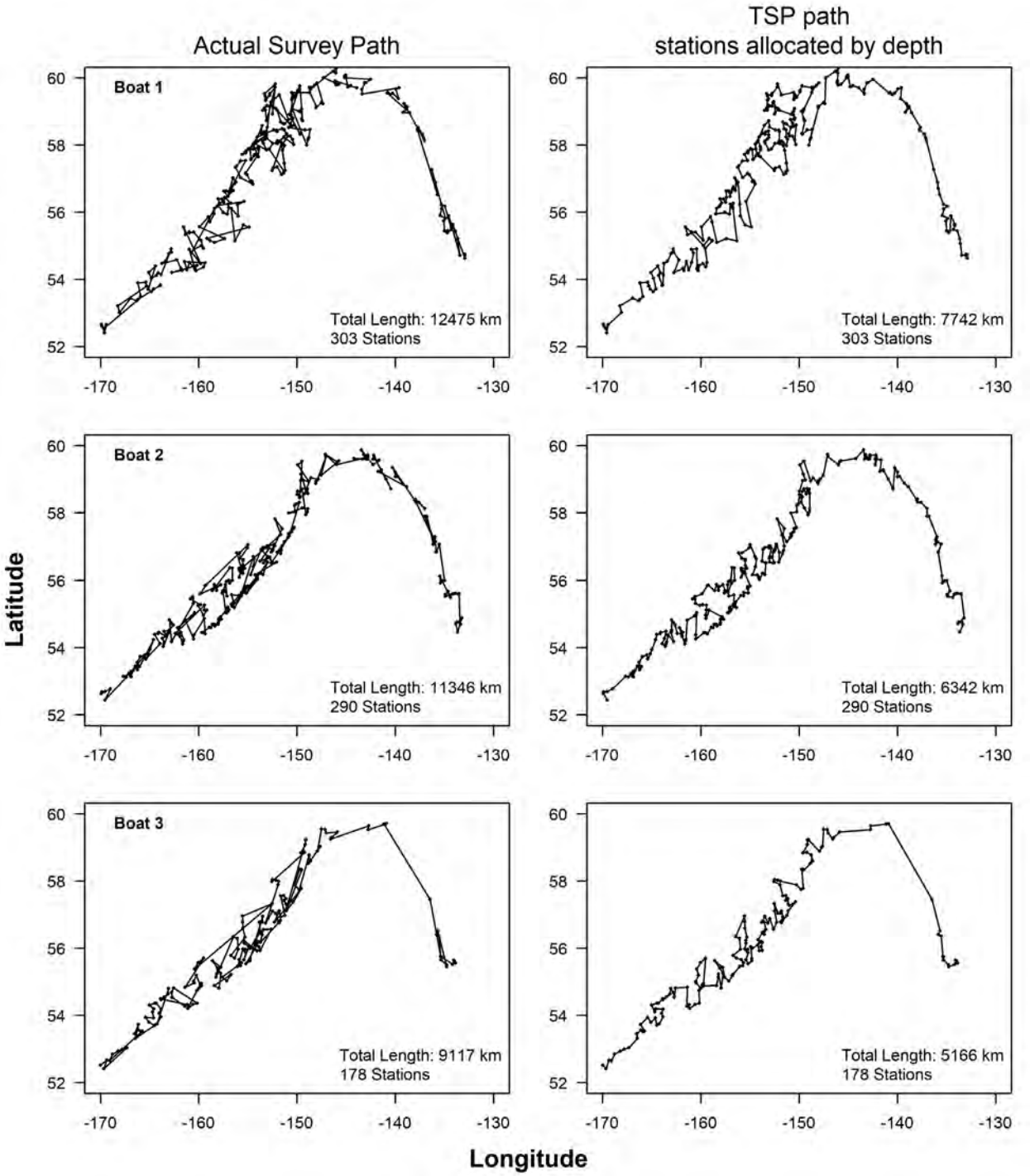
Appendix Figure D-7. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2011. Surveys start at the most western station and traverse eastward.



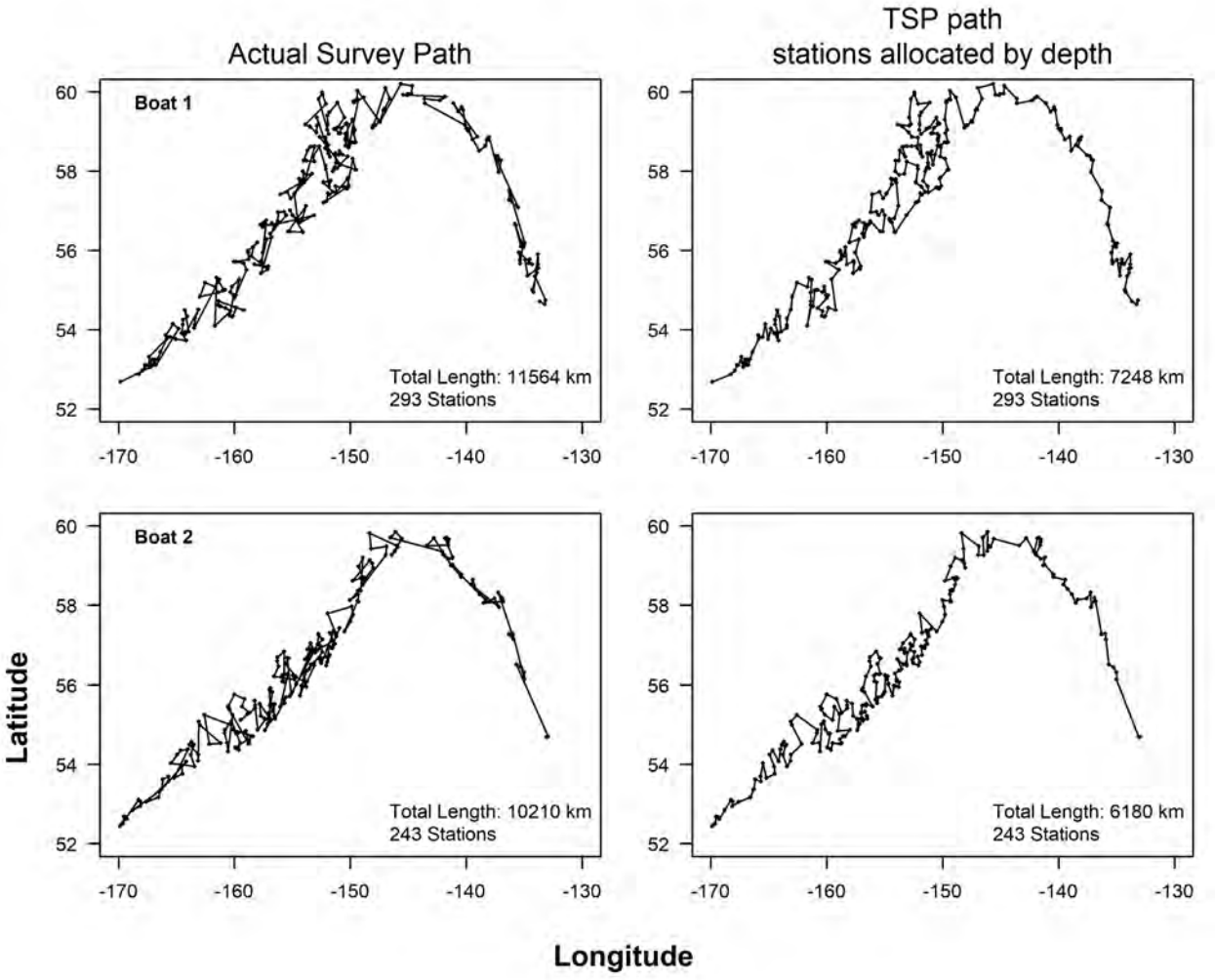
Appendix Figure D-8. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2013. Surveys start at the most western station and traverse eastward.



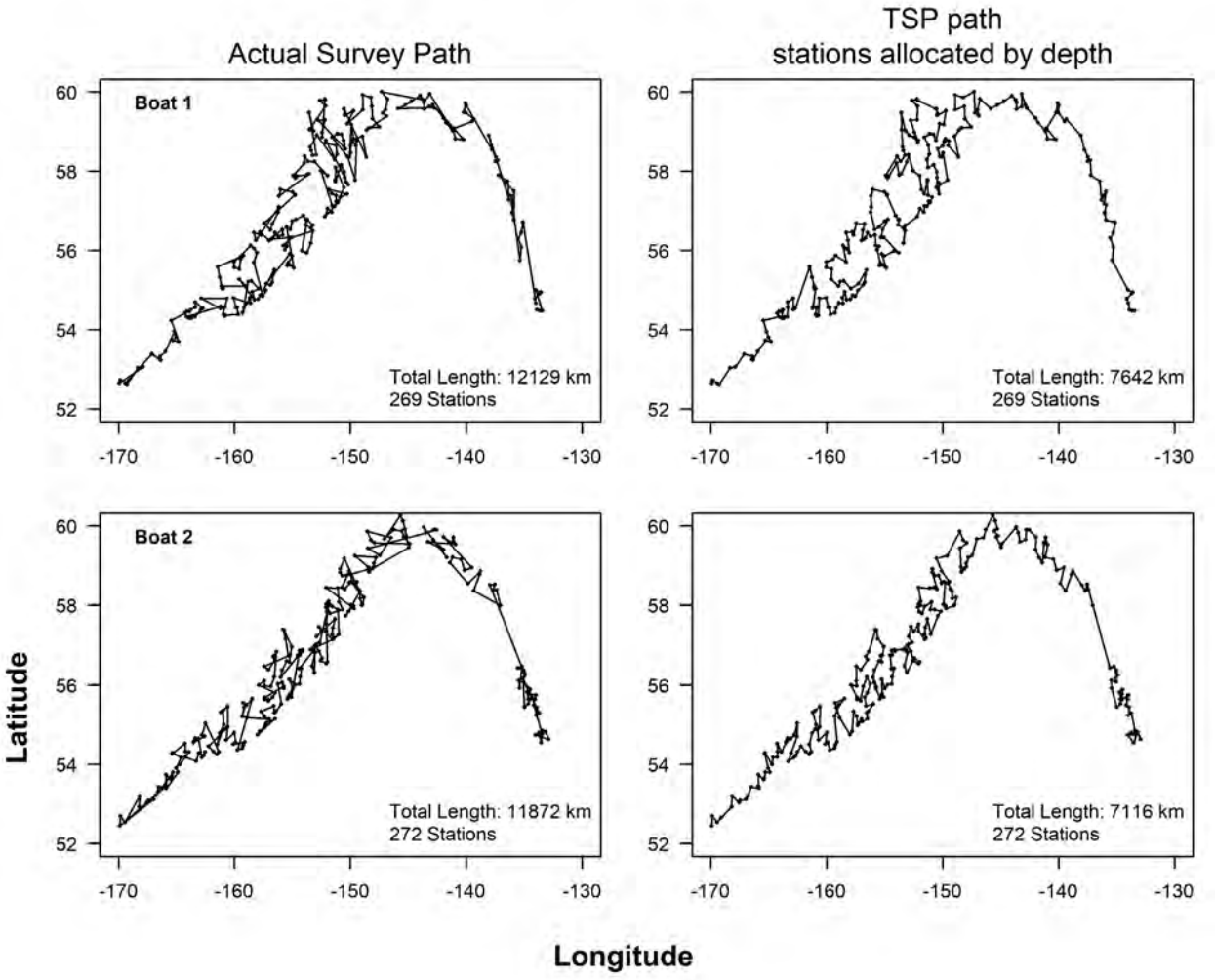
Appendix Figure D-9. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2015. Surveys start at the most western station and traverse eastward.



Appendix Figure D-10. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2017. Surveys start at the most western station and traverse eastward.



Appendix Figure D-11. -- Survey paths observed by the survey (left) for each boat (rows) along with the optimal shortest path from solving the Travelling Salesperson Problem (right) using the same stations allocated to each boat in 2019. Surveys start at the most western station and traverse eastward.



U.S. Secretary of Commerce
Gina M. Raimondo

Under Secretary of Commerce for Oceans and Atmosphere
Dr. Richard W. Spinrad

Assistant Administrator, National Marine Fisheries Service. Also serving as Acting Assistant Secretary of Commerce for Oceans and Atmosphere, and Deputy NOAA Administrator
Janet Coit

March 2022

www.nmfs.noaa.gov

OFFICIAL BUSINESS

National Marine Fisheries Service
Alaska Fisheries Science Center
7600 Sand Point Way N.E.
Seattle, WA 98115-6349