National Marine
Fisheries Service
NOAA

Fishery Bulletin
established in 1881

Spencer F. Baird
First U.S. Commissioner
of Fisheries and founder
of *Fishery Bulletin*

Effect of reduction in spatial survey effort on indices of bottom temperature for the eastern Bering Sea

Cynthia Yeung
Stan Kotwicki
Sean K. Rohan

Email address for contact author: cynthia.yeung@noaa.gov

Resource Assessment and Conservation Engineering Division
Alaska Fisheries Science Center
National Marine Fisheries Service, NOAA
7600 Sand Point Way NE
Seattle, Washington 98115

Abstract—Bottom temperature is routinely measured as part of the bottom-trawl survey conducted every summer on the continental shelf of the eastern Bering Sea by the NOAA Alaska Fisheries Science Center. These data are widely used in ecosystem, stock assessment, and ocean modeling. We assessed the effect of alternative sampling designs and effort reduction on the quality of bottom-temperature information from the survey. Simple-random and stratified-random sampling were simulated and compared with the systematic sampling of fixed stations in the regular grid used in the standard survey, with respect to the use of survey data in the estimation of bottom temperatures and related indices. The effort simulated ranged from 34% to 100% of the full effort. In the simulated surveys, the use of each of the 3 sampling designs resulted in values of bottom-temperature metrics that are close to those from the real survey, even with as little as half the effort. Lower effort resulted in larger and more variable prediction errors of the indices. The decrease in prediction performance is most noticeable at the 34% effort level. Systematic sampling performed slightly better than simple-random and stratified-random sampling. One reason for this difference in performance is that random sampling may have been less effective than the standard sampling in capturing a small cold pool that is characteristic of the current warm ocean state.

Manuscript submitted 17 January 2023.
Manuscript accepted 1 August 2023.
Fish. Bull. 121:112–123 (2023).
Online publication date: 28 August 2023.
doi: 10.7755/FB.121.3.5

All physiological rates of aquatic invertebrate and teleost fish species depend on body temperatures (Brown et al., 2004). Therefore, temperature is the principal driver of aquatic ecosystem dynamics (Kooijman, 2000; Poloczanska et al., 2016). The continental shelf of the eastern Bering Sea (EBS) is one of the most productive ecosystems in the world (Hood and Calder, 1981), supporting major groundfish fisheries, such as those that target walleye pollock (*Gadus chalcogrammus*) (Loughlin and Ohtani, 1999). Bottom water temperature is a key indicator of ecosystem state because of its pivotal role in the spatial distribution of groundfish species (Eisner et al., 2020; Kotwicki and Lauth, 2013), predator–prey interactions (Grüss et al., 2021), and demographic rates (Cooper et al., 2020).

The bottom temperature of the EBS is determined by an interplay between winter cooling, mixing due to wind and tide, and the spatial extent and persistence of ice cover (Sullivan et al., 2014). Bottom temperature is used to map the *cold pool*—a term typically used to refer to the layer of cold (<2°C) bottom water that forms below the pycnocline when sea ice freezes in the winter and extends southward over the middle of the shelf (at depths of 50–100 m) in summer in the EBS (Wyllie-Echeverria and Wooster, 1998; Stabeno et al., 2012). The size of the cold pool depends on the extent of sea ice during the previous winter. Under climate change, the area and thickness of sea ice are diminishing; ice forms later in the fall and retreats earlier in the spring (Overland et al., 2018; Stabeno and Bell, 2019). Since a warm thermal stanza across the Bering Sea began in late 2013 (NPFMC, 2016), there has been a series of record-breaking high water temperatures (Stabeno et al., 2019), and the cold pool has retracted northward, to the point of disappearing almost entirely from the EBS in 2018 (Rohan et al., 2022).

Bottom temperature in the EBS and cold pool indices are significant climate change indicators (Mueter and Litzow, 2008; Kotwicki and Lauth, 2013; Hollowed et al., 2020;) and environmental covariates in ecosystem and groundfish stock assessment (Thorson, 2019; Kearney et al., 2020; Rooper et al., 2021). In a cursory search in Google Scholar (website, accessed January 2023), about 8320 articles published since 2010 contained the phrase *eastern Bering Sea*, and in 10% and

7% of those articles *bottom temperature* and *cold pool* were also mentioned, respectively.

The primary source of data on summer bottom temperature for the EBS shelf is the bottom-trawl survey conducted every summer (June–August) since 1971 by the NOAA Alaska Fisheries Science Center for the assessment of groundfish stocks (Armistead and Nichol, 1993). The survey area includes the entire continental shelf at the depths of 20–200 m, and a regular grid of fixed stations is used for systematic sampling (Fig. 1). Temperature has been routinely recorded with each trawl sample in the standard survey since 1982. Survey temperature data are used to calibrate and validate the Bering10K Modeling Suite (Capotondi et al., 2019; Kearney et al., 2020). Although the cold pool is most commonly defined as bottom water colder than 2°C, 0°C and 1°C have also been used in the study of the spatial dynamics of fish populations (Kotwicki and Lauth, 2013; Thorson et al., 2017; Nichol et al., 2019).

Long-term ecological studies and resource surveys are expensive and complex to conduct. Worldwide, these programs are vulnerable to reductions in effort due to financial, logistical, or natural obstacles (ICES, 2020). The increasing disparity between funding allocation for the NOAA bottom-trawl survey and rising costs is also a reality that prompts considerations of effort reduction. Alternative sampling strategies under the constraint of reduced effort are often simulated to evaluate how the quality of survey data would be affected and whether program objectives could be achieved (e.g., Liu et al., 2011; Del Vecchio et al., 2019). In this study, we simulated surveys with systematic, simple-random, or stratified-random sampling and with different levels of effort, measured as the number of stations sampled, and compared estimation of bottom temperatures and related indices based on values from those simulated surveys with estimation based on data from the real standard surveys for which systematic sampling of fixed stations in a regular grid was used. The effort levels simulated ranged from 34% to 100% of the full effort of sampling 350 stations.

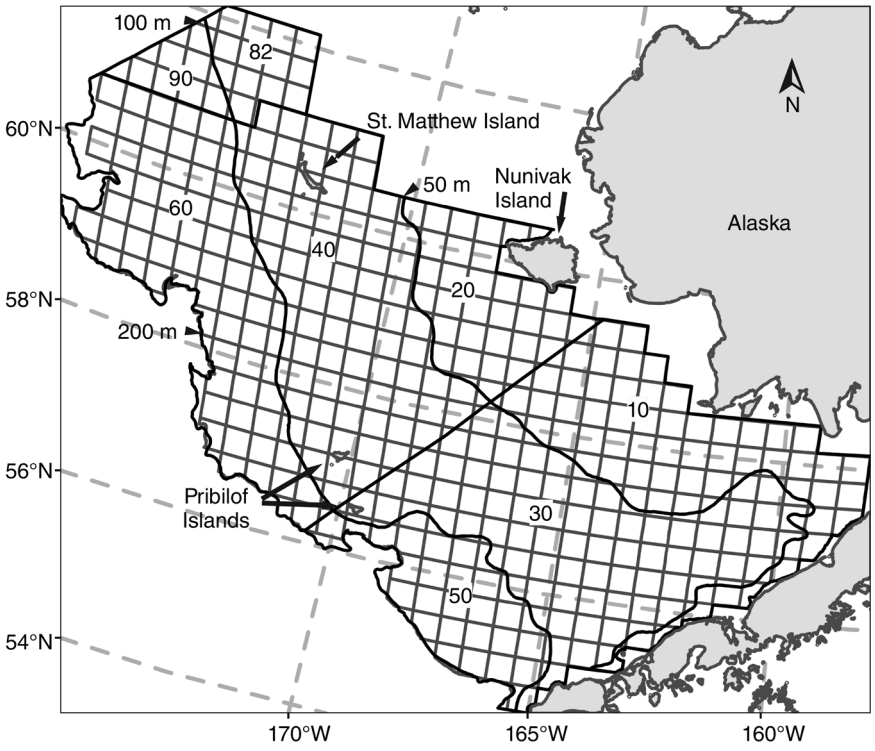

Figure 1

Map showing the regular grid used in the systematic sampling of the bottom-trawl survey conducted every summer on the continental shelf of the eastern Bering Sea by the NOAA Alaska Fisheries Science Center and used in the simulations of surveys in this study. Trawl samples are collected at stations located approximately at the center of each cell, each 20 × 20 nautical miles (37 km^2), in the grid. The survey area is divided into strata based on latitude and depth (strata are outlined with black lines and labeled with their assigned numbers), with the offshore boundaries of strata aligned with the 50-m, 100-m, and 200-m isobaths. The major islands on the shelf are indicated.

Materials and methods

Observations of bottom temperature

The bottom-trawl survey on the continental shelf of the EBS (hereafter referred to as *the survey*) has a systematic, regular-grid sampling design consisting of 350 cells of 20 × 20 nautical miles (nmi) (37.0 × 37.0 km) each (Fig. 1). The grid is divided into several strata that correspond to a combination of latitudinal and depth-associated biophysical domains that fall in 3 groups: inner shelf (depths <50 m), middle shelf (depths of 50–100 m), and outer shelf (depths >100 and ≤200 m) (Kinder and Schumacher, 1981; Coachman, 1986). Each stratum is assigned a number so that they can be referenced for management purposes (Fig. 1). Sampling during the survey progresses through the grid from the southeast corner to the northwest corner over an average of 57 days from late May to early August (Rohan et al., 2022). A standard trawl sample is taken from a location (or *station*) at approximately the center of each cell. The station configuration currently used for the survey was adopted in 1987 (Conner and Lauth, 2017). Bottom temperature (hereafter referred to as *temperature*) is recorded at each station. The various instruments used over the years, including an expendable bathythermograph and a digital temperature-depth recorder, are detailed in Rohan et al. (2022). Since 1996, the instrument for recording temperature has been attached to the head rope of the trawl net.

A spatial map of summer temperatures in the Bering Sea is produced annually as part of the survey and is a composite of observations from the survey duration of more than 2 months (Rohan et al., 2022). In this study, for each year from 1987 through 2019, temperature was simulated for the subset of stations that was considered the full effort: the 350 stations (Fig. 1) at the centers of cells, excluding the 26 stations located at the corners of cells around St. Matthew Island and the Pribilof Islands that are used in assessment of local populations of blue king crab (*Paralithodes platypus*) (Armistead and Nichol, 1993).

Simulated surveys of temperature

Spatial analyses were conducted by using popular spatial packages in the R statistical computing environment (vers. 3.6.3; R Core Team, 2020): gstat (vers. 2.1-1; Gräler et al., 2016), raster (vers. 3.3-13; Hijmans, 2020), sf (vers. 1.0-12; Pebesma, 2018), and sp (vers. 1.6-0; Pebesma and Bivand, 2005).

A regular grid of 494 rows and 565 columns (279,110 pixels) and with a 2-km resolution was superimposed over the survey area. Observed temperature at the standard set of 350 stations was interpolated over the grid by using ordinary kriging with Stein's Matérn autocovariance model (Stein, 1999), as implemented in the R package gstat, to produce annual temperature spatial surfaces (*rasters*). This process is the standard method for calculating temperature data products with information from the bottom-trawl survey in the Bering Sea (Rohan et al., 2022). The Matérn model was fitted to the sample variogram (smoothness=0.5; the sill and range automatically fitted by using gstat; Gräler et al., 2016). The variogram was used to interpolate pixel values over the raster. Only the pixels that partially or entirely overlapped the survey area were assigned values to interpolate the raster (R):

$$R = \left\{r_1\right\}_{l=1}^{N_r=123,274}, \quad (1)$$

where r_1 = pixel l, and
N_r = the total number of pixels in R.

The raster produced from observed temperatures was defined as the "true" thermal state of the EBS shelf.

Several combinations of sampling design and effort were applied to resample the raster based on observed temperatures of each year in the set of years (Y) to simulate surveys:

$$Y = \{1987, 1988, \ldots, 2019\} = \left\{y_i\right\}_{i=1}^{N_y=33}, \quad (2)$$

where y_i = year i, and
N_y = the total number of years in Y.

The following sampling designs were used: 1) systematic, regular-grid (also referred to as *regular*), 2) simple-random (also referred to as *random*), and 3) stratified-random with proportionate allocation by area of stratum (also referred to as *stratified*) (Fig. 1). For random and stratified sampling, 4 effort levels (E), equivalent to 34%, 46%, 66%, and 100% of the full effort, were simulated as follows:

$$E = \{120, 160, 230, 350 \text{ stations}\} = \left\{e_j\right\}_{j=1}^{N_e=4}, \quad (3)$$

where e_j = effort j, and
N_e = the total number of levels in E.

The 3 reduced regular sampling efforts that were simulated were based on grid cell size: 35 nmi^2 (64.0 km^2) for the effort of 120 stations, 30 nmi^2 (55.6 km^2) for the effort of 163 stations, and 25 nmi^2 (46.3 km^2) for the effort of 226 stations. The full regular sampling effort of 350 stations is that of the actual survey with the cell size of 20 nmi^2 (37.0 km^2). The expected number of stations in the simulated set with regular effort is practically equivalent to the sets with random and stratified effort based on the exact number of stations.

For each combination of sampling design and effort, 100 replicate sets of different station locations were drawn from the survey grid with the spsample function in the R package sp (Bivand et al., 2013), with the replicate sets (S) defined as follows:

$$S = \left\{s_k\right\}_{k=1}^{N_s=100}, \quad (4)$$

where s_k = replicate set k, and
N_s = the total number of replicate sets in S.

Each replicate set represents a simulated survey. Temperature values were extracted (predicted temperature) at each

station from the raster based on observed temperatures. The point values of predicted temperature are used to interpolate a predicted temperature raster by using the same aforementioned methods for generating an observed temperature raster. The steps of the simulation are summarized schematically in Figure 2.

Performance evaluation

The performance of the simulated surveys in replicating temperatures from the real annual surveys was evaluated on the basis of the bias (ε) and the root mean squared error (RMSE) of each of the predicted temperature indices by using these general equations:

$$\varepsilon = \frac{1}{N}\sum_{t=1}^{N}\left(\hat{x}_t - x_t\right), \text{ and} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{x}_t - x_t\right)^2}, \quad (6)$$

where $\hat{x}_t$ = the predicted temperature index for t=1, . . .,N, where the total number of samples, N, depends on the index.

The term *error* will hereafter refer generally to both bias and RMSE. The fundamental unit is the temperature value in each pixel of a raster, from which temperature indices are constructed. We examined the following indices: mean temperatures of the whole shelf (*shelf*) and shelf stratum (*stratum*), total cold pool area, the offshore (western-most longitude) and southern (southern-most latitude) extents of the cold pool, and the depth at the offshore extent of the cold pool. In addition to the common definition of the cold pool as bottom water with temperature <2°C (CP2), definitions with temperature thresholds of <1°C (CP1) and <0°C (CP0) were also evaluated. These indices comprehensively describe the general thermal state of the shelf and the location of the cold pool for a given year for the purposes of ecosystem and fishery management (Nichol et al., 2019; Siddon, 2021; Thorson et al., 2017; Uchiyama et al., 2020).

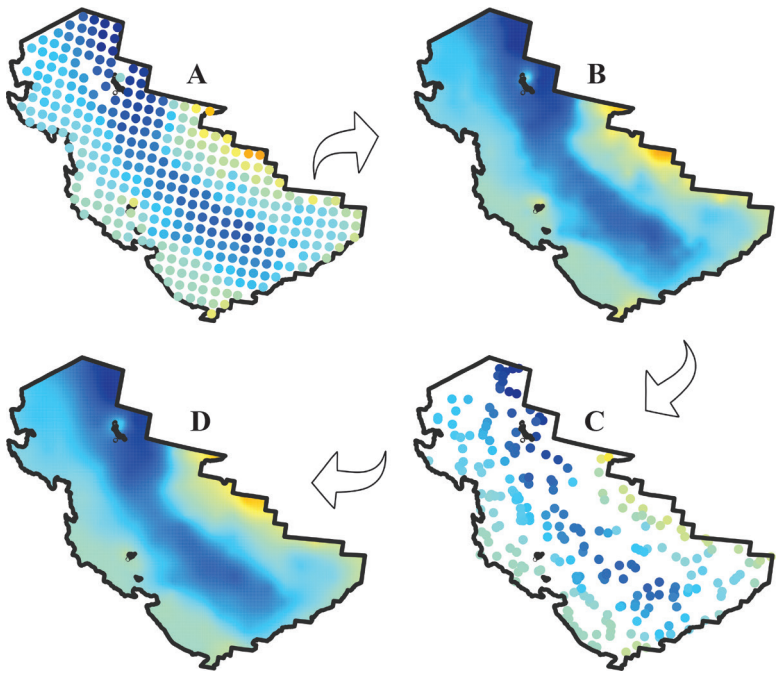

Figure 2

Schematic of the 4 main steps used to simulate a survey of bottom temperature on the continental shelf of the eastern Bering Sea. The steps are as follows: (**A**) start with temperatures observed at 350 regularly spaced stations of the real bottom-trawl survey from 1987 through 2019 (points indicate stations), (**B**) interpolate point values to generate a raster surface for observed temperature, (**C**) superimpose on the observed temperature raster stations simulated by using 1 of 3 sampling designs (regular, simple random [shown here], or stratified random) and 1 of 4 effort levels (number of stations: 120, 160, 230, or 350) and extract predicted temperatures at those stations, and (**D**) interpolate point values to generate a raster surface for predicted temperature. Values for corresponding pixels (simulated stations) in the rasters of predicted and observed temperature and related temperature indices were compared to evaluate prediction performance of the sampling designs used to simulate surveys.

Temperature rasters

To evaluate the prediction performance of a single raster, the error was calculated for each pixel l, and then the mean pixel error of the raster was determined across all pixels of the raster (N=123,274). To evaluate the variability among replicate rasters of a given combination of effort and year (or *treatment*) for each sampling design, the mean error of pixel l was calculated across the 100 replicates by using these equations:

$$\varepsilon_{ij.l} = \frac{1}{N_s}\sum_{k=1}^{N_s}\varepsilon_{kl}, \quad (7)$$

where $\varepsilon_{ij.l}$ = mean bias of pixel l for year i and effort j, and
ε_{kl} = mean bias of pixel l across all k replicate sets; and

$$RMSE_{ij.l} = \sqrt{\frac{1}{N_s}\sum_{k=1}^{N_s}\left(\varepsilon_{kl}\right)^2}, \quad (8)$$

where $RMSE_{ij.l}$ = mean RMSE of pixel l for year i and effort j.

The overall mean error of a given design and effort combination was averaged

across all years as follows ($N=N_y \times N_r=33 \times 123{,}274=4{,}068{,}042$):

$$\varepsilon_{j..} = \frac{1}{N_y N_r} \sum_{i=1}^{N_y} \sum_{l=1}^{N_r} \varepsilon_{ij.l}, \text{and} \quad (9)$$

$$RMSE_{j..} = \frac{1}{N_y N_r} \sum_{i=1}^{N_y} \sum_{l=1}^{N_r} RMSE_{ij.l}. \quad (10)$$

The standard deviation (SD) of the overall bias was determined with this equation:

$$SD(\varepsilon) = SD(\varepsilon_{ij..}) = \sqrt{\frac{1}{N_y N_r - 1} \sum_{i=1}^{N_y} \sum_{l=1}^{N_r} \left(\varepsilon_{ij.l} - \varepsilon_{j..}\right)^2}. \quad (11)$$

Mean temperatures of shelf and stratum

Predicted and observed annual mean temperatures of the whole shelf and individual stratum in the EBS were calculated by averaging all pixelated temperature values within the unit (shelf or stratum) in each raster. The error in annual mean shelf or stratum temperature was evaluated between each replicate raster based on predicted or observed temperatures.

Cold pool indices

The area of the cold pool was calculated as the number of pixels of the cold pool multiplied by the pixel size (2 km^2), for each of the 3 definitions of cold pool based on different boundaries (temperature thresholds) of the cold pool (0°C, 1°C, or 2°C). The error of the area was also expressed as a percentage of the observed area to evaluate their correlation. The offshore and southern extents of the cold pool are defined as the western-most longitude, and southern-most latitude of the cold pool identified on the raster. The depth of the cold pool at the geographic coordinates of its offshore extent was extracted from a bathymetry raster of the entire part of the U.S. exclusive economic zone off Alaska. The original raster with a 20-m resolution was compiled from 18.6 billion bathymetric soundings obtained from the NOAA National Center for Environmental Information (Lewis[1]). For this analysis, the raster was resampled to a 2-km resolution and clipped to the survey grid by using the Database Management Toolbox for raster processing in ArcMap[2], part of ArcGIS Desktop, vers. 10.7. (Esri, Redlands, CA).

Results

Within a sampling design, lower effort resulted in larger and more variable prediction errors in temperature metrics. Only exceptions will be mentioned hereafter.

[1] Lewis, S. 2018. Personal commun. Sustain. Fish. Div., Alsk. Reg. Off., Natl. Mar. Fish. Serv., NOAA, P.O. Box 21668, Juneau, AK 99802-1668.

[2] Mention of trade names or commercial companies is for identification purposes only and does not imply endorsement by the National Marine Fisheries Service, NOAA.

Temperature rasters

Prediction errors of individual rasters and among replicate sets of rasters were both low regardless of the combination of sampling design and effort. The overall mean values of pixel bias of the rasters for all combinations were ≤0.01°C (SD <0.3), and the overall mean values of pixel RMSE of the rasters for all combinations were <0.3°C (SD <0.2) (Table 1). The difference in prediction performance among designs for the same effort was small. Random sampling resulted in slightly higher RMSE values by magnitudes of tenths of degrees Celsius. Replicate rasters of predicted temperature within a treatment were similar to each other and to the corresponding raster of observed temperature. The similarity between annual observed temperature rasters (Fig. 3A) and mean predicted temperature rasters (N_s=100) from the survey simulation with random sampling and an effort of 230 stations (Fig. 3B) exemplifies the close visual resemblance of temperature rasters regardless of design and effort combination.

Mean temperatures

Shelf The means for bias in temperature on the entire shelf for all treatments were ≤±0.02°C (SD ≤0.06). The RMSE values in shelf temperatures for all treatments were <0.4°C. In almost all cases, predicted annual mean

Table 1

Overall mean bias ($\bar{\varepsilon}$) and overall root mean squared error (RMSE) and the standard deviation (SD) of the overall mean bias in the prediction of bottom temperature on the continental shelf of the eastern Bering Sea in rasters created by simulating bottom-trawl surveys for the years from 1987 through 2019. The statistics summarize the variability among replicates within a given combination of sampling design and survey effort (N=123,274 pixels per mean annual raster×33 years=4,068,042). The 3 sampling designs used in simulations were systematic (regular), simple random (random), and stratified random (stratified). Effort is given as the number of stations sampled to interpolate a raster of predicted temperature.

		Bias		
Design	Effort	$\bar{\varepsilon}$	SD	RMSE
Regular	120	−0.011	0.230	0.195
Regular	160	−0.008	0.190	0.155
Regular	230	−0.006	0.147	0.113
Random	120	−0.010	0.285	0.291
Random	160	−0.008	0.243	0.244
Random	230	−0.006	0.195	0.192
Random	350	−0.004	0.147	0.141
Stratified	120	−0.011	0.253	0.237
Stratified	160	−0.009	0.212	0.193
Stratified	230	−0.006	0.165	0.147
Stratified	350	−0.004	0.121	0.103

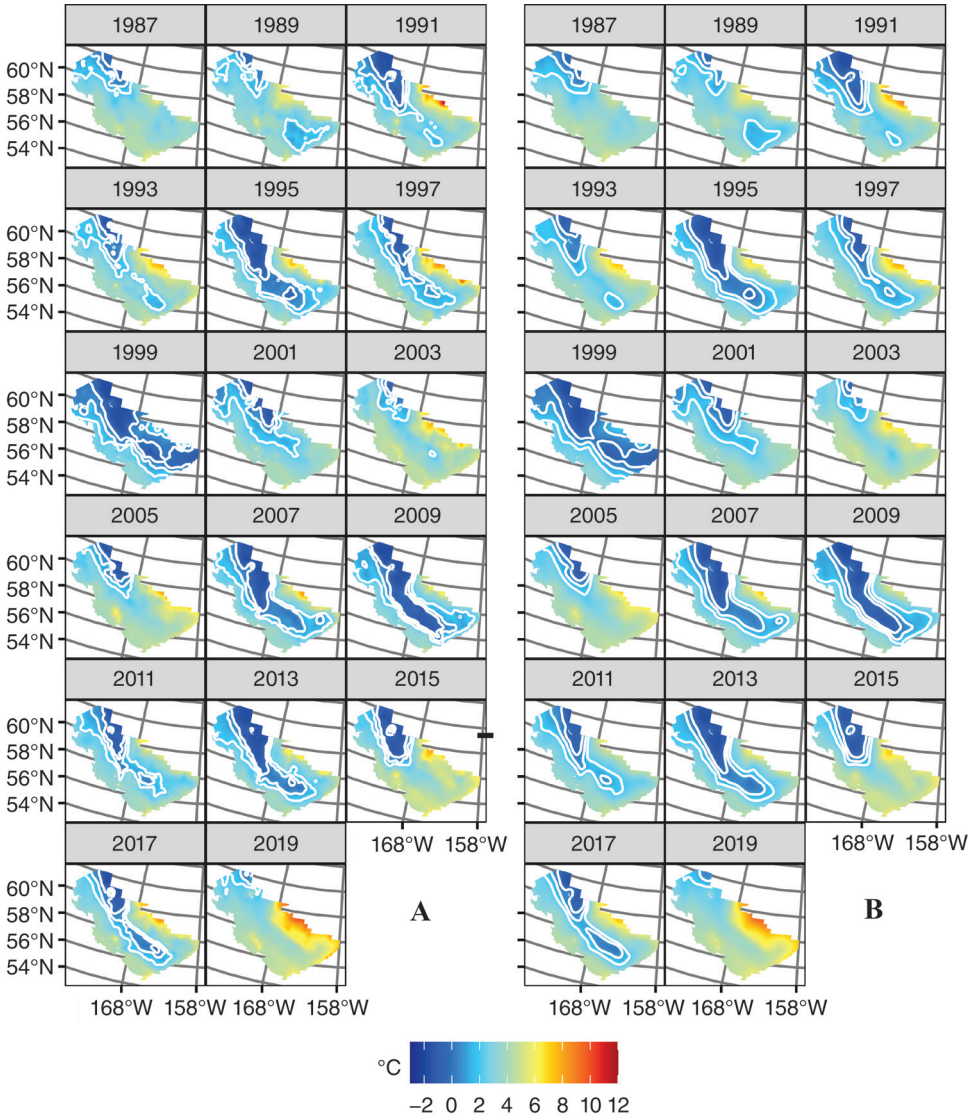

Figure 3

Annual bottom temperature in the eastern Bering Sea for the odd-numbered years from 1987 through 2019: (**A**) rasters of interpolated values based on observations recorded during actual bottom-trawl surveys and (**B**) rasters of simulated values based on the means of 100 replicate rasters produced by simulating random sampling of 230 stations superimposed on the observed temperature raster. The white lines indicate the 0°C, 1°C, and 2°C isotherms.

temperatures were lower than observed temperatures (negative bias). At the lowest effort of 120 stations, random and stratified sampling resulted in errors with greater interannual variability. This variability was more obvious in the bias values, which for some years were of smaller magnitude (less negative) than bias values produced in treatments with higher effort (Fig. 4). The values of RMSE (not shown in a figure) were consistently higher and more variable for treatments with lower effort levels but otherwise had similar patterns across treatments.

Stratum Predicted mean stratum temperatures were higher than observed temperatures (positive bias) in the middle-shelf strata 30, 40, and 82 (for locations of numbered strata, see Figure 1), the principal domain of the cold pool, but lower than observed temperatures in all other strata (Suppl. Fig. 1). The RMSE values were generally around 0.2°C. They were lowest for outer-shelf stratum 60 and relatively higher and more variable for the inner-shelf strata 10 and 20 and in middle-shelf stratum 82. The RMSE for stratum 50 was anomalously high in 1992 (Suppl. Fig. 2), a difference that was likely

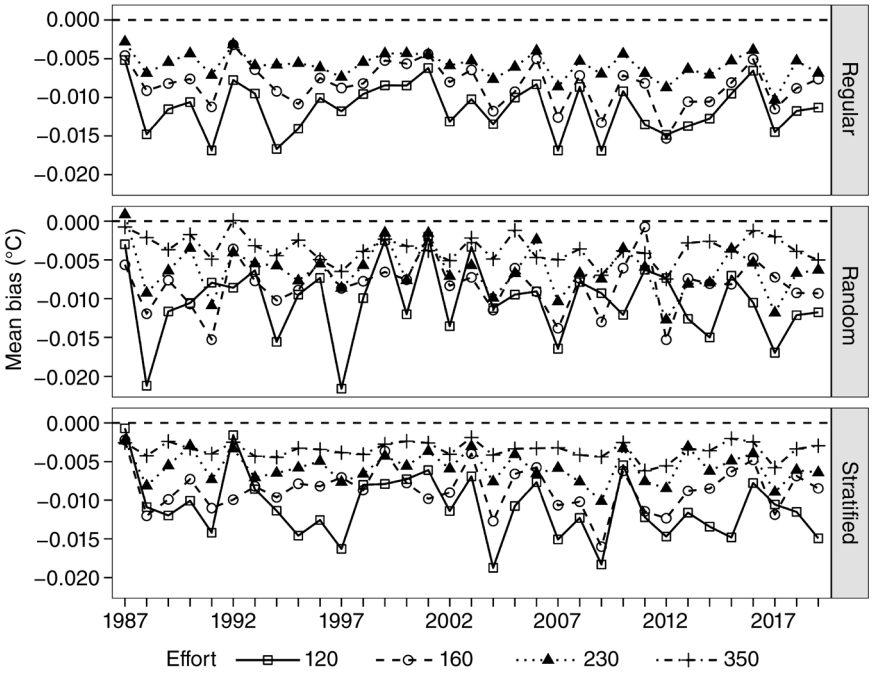

Figure 4

Mean bias in prediction of the mean bottom temperature of the entire continental shelf in the eastern Bering Sea, by sampling design (regular, simple random, or stratified random) and effort level (number of stations) for each survey year from 1987 through 2019.

associated with artifacts of interpolation (fragments of water in stratum 50 with temperatures <2°C in the raster based on observed temperatures that were not replicated in the rasters based on predicted temperatures; Fig. 3). Random sampling resulted in errors that are relatively larger than errors from the use of other designs at the same effort. The difference in level of errors was most noticeable for strata 20 and 82 at the effort of 120 stations.

Cold pool indices

Performance was similar among sampling designs, in predicting total cold pool area, the offshore and southern extents of the cold pool, and the depth at the offshore extent. Regular sampling generally resulted in slightly smaller errors than random and stratified sampling. Differences were in the magnitude of tenths of a unit.

Area Prediction bias in the cold pool area as a percentage of the observed cold pool area was small for CP2, except in the simulation for 2018 (Fig. 5), when bottom water colder than 2°C occupied only 1.2% of the survey area along the northern boundary and water colder than 1°C was not observed (Figs. 3A and 6). The RMSE as a percentage of the observed cold pool area had very similar patterns to the bias percentage. The bias and RMSE percentages were highly correlated, with coefficient of correlations (r) ranging from −0.8, for both CP1 and CP2, to approximately 1 for CP0 (P<0.001, N=352). Excluding the percentage for 2018, the bias percentages across all treatments were larger and more variable for the colder boundaries of the cold pool (0°C and 1°C) and skewed more toward the underestimation of the area: the range for CP2 was −10–4%, the range for CP1 was −30–12%, and the range for CP0 was −97–1%. The bias percentages were moderately correlated with observed cold pool areas (CP2: r=0.37; CP1: r=0.26; CP0: r=0.44; P<0.001, N=11 treatments×33 years×100 replicates of simulated cold pool areas=36,300). In all instances, large negative biases were associated with small cold pool areas.

Extent and depth Bias and RMSE patterns for indices of cold pool location were largely similar. For any combination of sampling design and effort, mean biases in predicting cold pool extent were ≤0.4°N (SD ≤0.3) and ≤±0.2°W (SD ≤0.5); RMSE values were ≤0.6° for both latitude and longitude (for bias plots, see Supplementary Figures 3–5; plots for RMSE are not provided because the patterns are similar to those for bias). The magnitude of the errors was associated with the cold pool boundary and the survey year (i.e., the thermal environment) and rarely larger than a fraction of a degree in longitude or latitude. The predicted southern extent was farther north than the observed

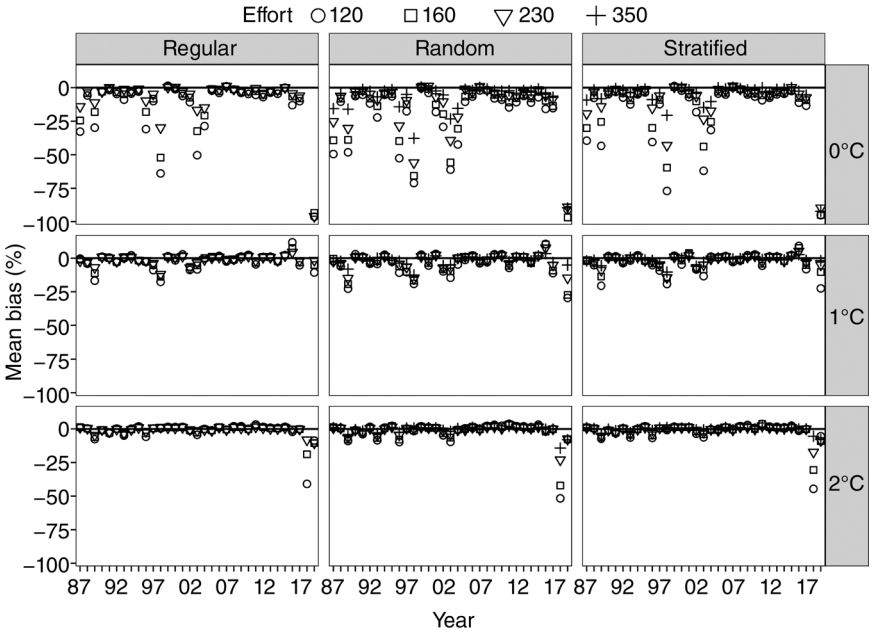

Figure 5

Mean bias in prediction of the cold pool area on the continental shelf in the eastern Bering Sea as a percentage of the observed cold pool area, by sampling design (regular, simple random, or stratified random) and effort level (number of stations) for each survey year from 1987 through 2019. The cold pool is a layer of cold (<0°C, <1°C, or <2°C) bottom water that forms on the shelf below the pycnocline.

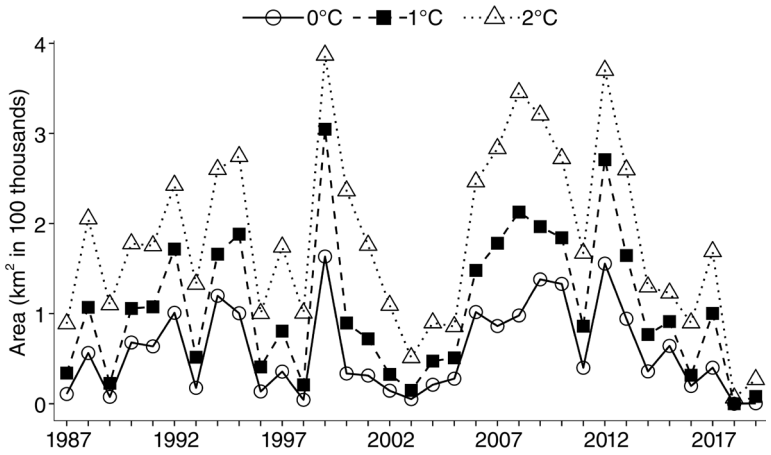

Figure 6

Annual observed area of the cold pool, a layer of cold (<0°C, <1°C, or <2°C) bottom water that forms below the pycnocline, on the continental shelf of the eastern Bering Sea from 1987 through 2019.

extent (positive bias) (Suppl. Fig. 3). The bias in predicting the southern extent of CP2 was largest for 2002 and 2003, for which observed temperature rasters show small, isolated pools of water <2°C farther south that were detached from the main pool, and for 2019, for which the main pool was tiny with an irregular southern boundary (Fig. 3A). The detached pools did not appear in the predicted temperature rasters for the corresponding years (Fig. 3B); hence, the predicted southern extent was farther north than the rasters based on observed temperatures, resulting in relatively large errors. The predicted offshore extent was mostly farther offshore than the observed extent for CP2 and CP1 (negative bias), but the trend was reversed for CP0, for which the biases were mostly positive (Suppl. Fig. 4).

Biases in predicting the depth at the offshore extent were larger (both mean and SD) for the warmer boundaries of the cold pool (Suppl. Fig. 5). For CP0 and CP1, the mean bias range was 1–4 m (RMSE: 1–4 m). For CP2, the mean bias ranged from −5 to −9 m (RMSE: 3–4 m) and was more variable annually than for CP0 and CP1. The largest bias of CP2, at about ±100 m, occurred in the simulations for 2002 and 2016 (Suppl. Fig. 5). The mean predicted offshore depth was shallower than the observed depth for CP0 and CP1 (positive bias) but deeper than the observed depth for CP2 (negative bias).

Discussion

In this study, simulated surveys with regular, simple-random, and stratified-random sampling of the continental shelf of the EBS at the same level of effort have comparable performance in predicting bottom temperatures and temperature-related indices. Regular sampling performed slightly better than the other designs; simple-random sampling performed about the same as stratified-random sampling. Lower effort resulted in larger prediction errors of the indices—on the scale of fractions of a degree for indices of mean temperature and cold pool location (latitude and longitude), thousands of square kilometers for cold pool area, and tens of meters in depth for offshore extent. Whether differences on that scale for an index is significant depends on the application. For example, the performance of ecosystem models and forecasting systems may depend on spatial resolution, coverage, and accuracy of the observational data (Capotondi et al., 2019; Thorson, 2019), whereas current qualitative assessment of ecosystem state and trends may be more tolerant of uncertainties (Zador et al., 2016). Interannual variability was most noticeable at the lowest effort of 120 stations, indicating that this level of effort may be a critical threshold for effort reduction.

The cold pool boundary is most commonly defined as water <2°C (CP2) in the literature. Kotwicki and Lauth (2013) suggested that water <1°C (CP1) better describes the temperature preferences of groundfish and crab species in the EBS. As the benthic environment becomes warmer and the cold pool becomes smaller under climate change, colder boundaries of the cold pool may be increasingly more difficult to delineate with high certainty, especially with reduced effort. It may be necessary to review current temperature-based indices and change them to indices that will more adequately depict thermal variability in the benthic environment in the EBS. In the scenario of an ice-free EBS, thermal variability may have less effect on ecosystem dynamics.

Spatial interpolation is less accurate and generates larger prediction errors generally when the number of data points is small and not uniformly distributed (Achilleos, 2011). These kinds of data insufficiency may have caused the apparent fragmentation of the cold pool in the observed temperature rasters and resulted in relatively large errors in the estimation of the southern extent of the cold pool (e.g., in the simulation for 2003). Higher sampling density in the vicinity where the cold pool generally resides (i.e., the middle shelf and the northern inner shelf) may be necessary to better delineate a small cold pool and estimate cold pool indices.

Prediction errors of the depth at the offshore extent of the cold pool are highly variable, and biases can be as large as ±100 m for CP2 and CP1. The magnitude of errors is small for CP0 because it is located nearshore, usually in depths ≤50 m. The accuracy of the estimated depth at a location is dependent on the resolution and accuracy of the bathymetry raster. The finding of relatively large prediction errors in the index for the depth at the offshore extent of the cold pool indicates that reduced sampling may lead to high uncertainty in the estimation of this index. The current methods of estimation may also be a causal factor. We recommend that this index be used with caution and that the use of other more robust indices may be preferable.

The results of this study indicate that the use of all 3 survey designs can reproduce temperatures and indices of the EBS shelf from the real bottom-trawl survey, for which sampling of 350 stations in a regular grid was done systematically, even with as little as half the effort. Reducing the number of stations that are sampled has definite financial and likely logistical benefits. Total on-station time can be reduced, although the total transit time to cover the survey area cannot necessarily be changed. The current survey requires 2 months to complete with 2 vessels (Conner and Lauth, 2017). Effort reduction may shorten the total survey time and enable the composition of a spatial temperature map that characterizes temperature for a more compressed time frame. It is not apparent that the use of any of the survey designs examined in this study delivers distinctly better performance in the construction of temperature spatial maps and indices or achieves greater economy and efficiency in survey execution. Regular sampling appears to perform slightly better and has been the standard sampling design for the bottom-trawl survey on the EBS shelf for over 30 years. Random sampling may be less effective in capturing a very small cold pool (e.g., the cold pool in 2018).

Changing the spatial extent of the sampling effort will likely also change the survey schedule and the day of the year on which certain stations are sampled. On the basis of Danielson et al. (2011) and Cokelet (2016), the surface mixed layer would warm by an estimate of 3.0°C/d. Analyses indicate that, if stations with depths <50 m in the current grid—which presumably would be fully mixed—are sampled approximately a week or more earlier than their mean sampling day of year, cold pool indices would produce average negative (i.e., cold) biases less than −0.2°C for those stations; conversely, sampling later than their mean sampling day of year would result in a positive (i.e., warm) bias (Rohan et al., 2022). The effect on the temperature indices of the temporal effort shift associated with spatial effort reduction would need to be assessed. Relationships used in stock assessments to correct for variation in density and availability would need to be revisited because they are largely predicated on assumptions of the stationarity of temperature indices.

Conclusions

Long-term ecological studies are imperative to the understanding of climate-scale processes and informing management policy (Sukhotin and Berger, 2013; Hughes et al., 2017; Harvey et al., 2020). The bottom-trawl survey on the EBS shelf provides an invaluable time series of fishery-independent biological data and environmental data essential for stock assessment and fisheries and ecosystem research under climate change (e.g., Holsman et al., 2016; Grüss et al., 2020). The collection of temperature data is

only one of many ancillary tasks of the survey. Evaluation of how changes in sampling design and reduction in effort will affect the accuracy and precision of stock assessment information is ongoing (e.g., Oyafuso et al., 2021).

The future direction of the survey in terms of design and effort will be driven mainly by stock and environmental assessment needs within fiscal constraints. When considering stratification and sample allocations for future surveys, it is important to consider not only species distribution but also the distribution and importance of the temperature indices needed for these assessments. Alternatively, the measurement of temperature during the survey may be independent of the bottom-trawl sampling design if instruments that are not attached to the trawl gear are used. Temperature profiling can be achieved within a reasonably short time at a station or possibly even while the vessel is underway. If measurement of temperature is decoupled from trawl sampling, it may not be necessary to decrease temperature samples by the same amount as trawl samples or restrict temperature sampling only to trawl stations.

Resumen

La temperatura del fondo se mide rutinariamente como parte del estudio de arrastre de fondo realizado cada verano en la plataforma continental del Mar de Bering oriental por el Centro de Ciencias Pesqueras de Alaska de la NOAA. Estos datos se utilizan ampliamente en la evaluación de ecosistemas, evaluación de poblaciones y modelación oceánica. Evaluamos el efecto de diseños de muestreo alternativos y de la reducción del esfuerzo, sobre la calidad de la información de la temperatura del fondo obtenida en la prospección. Se simuló y comparó el muestreo aleatorio simple y el muestreo aleatorio estratificado con el muestreo sistemático de estaciones fijas en una rejilla regular, en relación con el uso de datos de prospección en la estimación de las temperaturas del fondo marino y los índices relacionados. El esfuerzo simulado osciló entre el 34% y el 100% del esfuerzo total. En las prospecciones simuladas, el uso de cada uno de los 3 diseños de muestreo resulto en valores de temperatura del fondo cercanos a los de la prospección real, incluso con tan solo la mitad del esfuerzo. A menor esfuerzo, los errores de predicción de los índices son mayores y más variables. La disminución de los resultados de predicción es más notable en el nivel de esfuerzo del 34%. Los resultados del muestreo sistemático son ligeramente mejores que los del muestreo aleatorio simple y el muestreo aleatorio estratificado. Una de las razones de esta diferencia es que el muestreo aleatorio pudo haber sido menos eficaz que el muestreo estándar a la hora de captar una pequeña onda fría característica del actual estado cálido del océano.

Acknowledgments

We thank all the people who worked on the bottom-trawl surveys in the Bering Sea through the years for the existence of the temperature data time series and the reviewers for helping us to improve this manuscript.

Literature cited

Achilleos, G. A.
2011. The inverse distance weighted interpolation method and error propagation mechanism—creating a DEM from an analogue topographical map. J. Spat. Sci. 56:283–304. Crossref

Armistead, C. E., and D. G. Nichol.
1993. 1990 bottom trawl survey of the eastern Bering Sea continental shelf. NOAA Tech. Memo. NMFS-AFSC-7, 190 p.

Bivand, R. S., E. Pebesma, and V. Gómez-Rubio.
2013. Applied spatial data analysis with R, 405 p. Springer, New York.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West.
2004. Toward a metabolic theory of ecology. Ecology 85:1771–1789. Crossref

Capotondi, A., M. Jacox, C. Bowler, M. Kavanaugh, P. Lehodey, D. Barrie, S. Brodie, S. Chaffron, W. Cheng, D. F. Dias, et al.
2019. Observational needs supporting marine ecosystems modeling and forecasting: from the global ocean to regional and coastal systems. Front. Mar. Sci. 6:00623. Crossref

Coachman, L. K.
1986. Circulation, water masses, and fluxes on the southeastern Bering Sea shelf. Cont. Shelf Res. 5:23–108. Crossref

Cokelet, E. D.
2016. 3-D water properties and geostrophic circulation on the eastern Bering Sea shelf. Deep-Sea Res., II 134:65–85. Crossref

Conner, J., and R. R. Lauth
2017. Results of the 2016 eastern Bering Sea continental shelf bottom trawl survey of groundfish and invertebrate resources. NOAA Tech. Memo. NMFS-AFSC-352, 159 p.

Cooper, D., L. A. Rogers, and T. Wilderbuer.
2020. Environmentally driven forecasts of northern rock sole (*Lepidopsetta polyxystra*) recruitment in the eastern Bering Sea. Fish. Oceanogr. 29:111–121. Crossref

Danielson, S., L. Eisner, T. Weingartner, and K. Aagaard.
2011. Thermal and haline variability over the central Bering Sea shelf: seasonal and interannual perspectives. Cont. Shelf Res. 31:539–554. Crossref

Del Vecchio, S., E. Fantinato, G. Silan, and G. Buffa.
2019. Trade-offs between sampling effort and data quality in habitat monitoring. Biodivers. Conserv. 28:55–73. Crossref

Eisner, L. B., Y. I. Zuenko, E. O. Basyuk, L. L. Britt, J. T. Duffy-Anderson, S. Kotwicki, C. Ladd, and W. Cheng.
2020. Environmental impacts on walleye pollock (*Gadus chalcogrammus*) distribution across the Bering Sea shelf. Deep-Sea Res., II 181–182:104881. Crossref

Gräler, B., E. Pebesma, and G. Heuvelink.
2016. Spatio-temporal interpolation using gstat. R J. 8(1):204–218.

Grüss, A., J. Gao, J. T. Thorson, C. N. Rooper, G. Thompson, J. L. Boldt, and R. Lauth.
2020. Estimating synchronous changes in condition and density in eastern Bering Sea fishes. Mar. Ecol. Prog. Ser. 635:169–185. Crossref

Grüss, A., J. T. Thorson, C. C. Stawitz, J. C. P. Reum, S. K. Rohan, and C. L. Barnes.
2021. Synthesis of interannual variability in spatial demographic processes supports the strong influence of cold-pool extent on eastern Bering Sea walleye pollock (*Gadus chalcogrammus*). Prog. Oceanogr. 194:102569. Crossref

Harvey, C. J., J. L. Fisher, J. F. Samhouri, G. D. Williams, T. B. Francis, K. C. Jacobson, Y. L. deReynier, M. E. Hunsicker, and N. Garfield.
2020. The importance of long-term ecological time series for integrated ecosystem assessment and ecosystem-based management. Prog. Oceanogr. 188:102418. Crossref

Hijmans, R. J.
2020. Raster: geographic data analysis and modeling. R package, vers. 3.3–13. [Available from website, accessed July 2020.]

Hollowed, A. B., K. K. Holsman, A. C. Haynie, A. J. Hermann, A. E. Punt, K. Aydin, J. N. Ianelli, S. Kasperski, W. Cheng, A. Faig, et al.
2020. Integrated modeling to evaluate climate change impacts on coupled social-ecological systems in Alaska. Front. Mar. Sci. 6:00775. Crossref

Holsman, K. K., J. Ianelli, K. Aydin, A. E. Punt, and E. A. Moffitt.
2016. A comparison of fisheries biological reference points estimated from temperature-specific multi-species and single-species climate-enhanced stock assessment models. Deep-Sea Res., II 134:360–378. Crossref

Hood, D. W., and J. A. Calder (eds.).
1981. The eastern Bering Sea shelf: oceanography and resources, vol. 1, 625 p. Off. Mar. Pollut. Assess., NOAA, Seattle, WA.

Hughes, B. B., R. Beas-Luna, A. K. Barner, K. Brewitt, D. R. Brumbaugh, E. B. Cerny-Chipman, S. L. Close, K. E. Coblentz, K. L. de Nesnera, S. T. Drobnitch, et al.
2017. Corrigendum: long-term studies contribute disproportionately to ecology and policy. BioScience 67:271–281. Crossref

ICES (International Council for the Exploration of the Sea).
2020. ICES workshop on unavoidable survey effort reduction (WKUSER). ICES Sci. Rep. vol. 2, issue 72, 92 p. [Available from website.]

Kearney, K., A. Hermann, W. Cheng, I. Ortiz, and K. Aydin.
2020. A coupled pelagic–benthic–sympagic biogeochemical model for the Bering Sea: documentation and validation of the BESTNPZ model (v2019.08.23) within a high-resolution regional ocean model. Geosci. Model Dev. 13:597–650. Crossref

Kinder, T. H., and J. D. Schumacher.
1981. Hydrographic structure over the continental shelf of the southeastern Bering Sea. *In* The eastern Bering Sea shelf: oceanography and resources, vol. 1 (D. W. Hood and J. A. Calder, eds.), p. 31–52. Off. Mar. Pollut. Assess., NOAA, Seattle, WA.

Kooijman, S. A. L. M.
2000. Dynamic energy and mass budgets in biological systems, 424 p. Cambridge Univ. Press, Cambridge, UK.

Kotwicki, S., and R. R. Lauth.
2013. Detecting temporal trends and environmentally-driven [sic] changes in the spatial distribution of bottom fishes and crabs on the eastern Bering Sea shelf. Deep-Sea Res., II 94:231–243. Crossref

Liu, Y., Y. Chen, J. Cheng, and J. Lu.
2011. An adaptive sampling method based on optimized sampling design for fishery-independent surveys with comparisons with conventional designs. Fish. Sci. 77:467–478. Crossref

Loughlin, T. R., and K. Ohtani (eds.).
1999. Dynamics of the Bering Sea. Alsk. Sea Grant Rep. AK-SG-99-03, 838 p. Univ. Alsk. Sea Grant, Fairbanks, AK.

Mueter, F. J., and M. A. Litzow.
2008. Sea ice retreat alters the biogeography of the Bering Sea continental shelf. Ecol. Appl. 18:309–320. Crossref

Nichol, D. G., S. Kotwicki, T. K. Wilderbuer, R. R. Lauth, and J. N. Ianelli.
2019. Availability of yellowfin sole *Limanda aspera* to the eastern Bering Sea trawl survey and its effect on estimates of survey biomass. Fish. Res. 211:319–330. Crossref

Overland, J. E., M. Wang, and T. J. Ballinger.
2018. Recent increased warming of the Alaskan marine Arctic due to midlatitude linkages. Adv. Atmos. Sci. 35:75–84. Crossref

Oyafuso, Z. S., L. A. K. Barnett, and S. Kotwicki.
2021. Incorporating spatiotemporal variability in multispecies survey design optimization addresses trade-offs in uncertainty. ICES J. Mar. Sci. 78:1288–1300. Crossref

Pebesma, E.
2018. Simple features for R: standardized support for spatial vector data. R J. 10(1):439–446. Crossref

Pebesma, E. J., and R. S. Bivand.
2005. Classes and methods for spatial data in R. R News 5(2):9–13.

Poloczanska, E. S., M. T. Burrows, C. J. Brown, J. García Molinos, B. S. Halpern, O. Hoegh-Guldberg, C. V. Kappel, P. J. Moore, A. J. Richardson, D. S. Schoeman, et al.
2016. Responses of marine organisms to climate change across oceans. Front. Mar. Sci. 3:00062. Crossref

R Core Team.
2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Available from website, accessed February 2020.]

Rohan, S. K., L. A. K. Barnett, and N. Charriere.
2022. Evaluating approaches to estimating mean temperature and cold pool area from Alaska Fisheries Science Center bottom trawl surveys of the eastern Bering Sea. NOAA Tech. Memo. NMFS-AFSC-456, 42 p.

Rooper, C. N., I. Ortiz, A. J. Hermann, N. Laman, W. Cheng, K. Kearney, and K. Aydin.
2021. Predicted shifts of groundfish distribution in the Eastern Bering Sea under climate change, with implications for fish populations and fisheries management. ICES J. Mar. Sci. 78:220–234. Crossref

Siddon, E. (ed.).
2021. Ecosystem status report 2021. Eastern Bering Sea, 249 p. North Pac. Fish. Manage. Council, Anchorage. AK. [Available from website.]

Stabeno, P. J., and S. W. Bell.
2019. Extreme conditions in the Bering Sea (2017–2018): record-breaking low sea-ice extent. Geophys. Res. Lett. 46:8952–8959. Crossref

Stabeno, P. J., N. B. Kachel, S. E. Moore, J. M. Napp, M. Sigler, A. Yamaguchi, and A. N. Zerbini.
2012. Comparison of warm and cold years on the southeastern Bering Sea shelf and some implications for the ecosystem. Deep-Sea Res., II 65–70:31–45. Crossref

Stabeno, P. J., R. L. Thoman, and K. Wood
2019. Recent warming in the Bering Sea and its impact on the ecosystem. *In* Arctic report card 2019 (Richter-Menge, J., M. L. Druckenmiller, and M. Jeffries, eds.), p. 81–87. [Available from website.]

Stein, M. L.
1999. Interpolation of spatial data: Some theory for kriging, 249 p. Springer, New York.

Sukhotin, A., and V. Berger.
2013. Long-term monitoring studies as a powerful tool in marine ecosystem research. Hydrobiologia 706:1–9.

Sullivan, M. E., N. B. Kachel, C. W. Mordy, S. A. Salo, and P. J. Stabeno.
2014. Sea ice and water column structure on the eastern Bering Sea shelf. Deep-Sea Res., II 109:39–56. Crossref

Thorson, J. T.
2019. 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. 64:2632–2645. Crossref

Thorson, J. T., J. N. Ianelli, and S. Kotwicki.
2017. The relative influence of temperature and size-structure on fish distribution shifts: a case-study on Walleye pollock in the Bering Sea. Fish Fish. 18:1073–1084. Crossref

Uchiyama, T., F. J. Mueter, and G. H. Kruse.
2020. Multispecies biomass dynamics models reveal effects of ocean temperature on predation of juvenile pollock in the eastern Bering Sea. Fish. Oceanogr. 29:10–22. Crossref

Wyllie-Echeverria, T., and W. S. Wooster.
1998. Year-to-year variations in Bering Sea ice cover and some consequences for fish distributions. Fish. Oceanogr. 7:159–170. Crossref

Zador, S., and E. Siddon (eds.).
2016. Ecosystem considerations 2016. Status of the eastern Bering Sea ecosystem, 205 p. North Pac. Fish. Manage. Council, Anchorage, AK. [Available from website.]

Zador, S. G., K. K. Holsman, K. Y. Aydin, and S. K. Gaichas.
2016. Ecosystem considerations in Alaska: the value of qualitative assessments. ICES J. Mar. Sci. 74:421–430. Crossref