Temperature data from the eastern Bering Sea continental shelf bottom trawl survey as used for hydrodynamic model validation and comparison

by

Kelly Kearney

Resource Ecology and Fisheries Management Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way N.E., Building 4
Seattle, Washington 98115
Affiliation: University of Washington, Cooperative Institute for Climate, Ocean and Ecosystem Studies (NOAA Alaska Fisheries Science)

March 2021

Abstract

Over the past four decades, temperature data have been collected across the eastern Bering Sea shelf as part of the annual bottom trawl survey conducted through the Alaska Fisheries Science Center. This dataset includes a spatially resolved annual time series of surface and bottom temperature, and serves as a primary observation-based temperature dataset against which regional ocean models of the region are validated. This report provides an overview of the data processing used to prepare the survey observations for model validation purposes. We then used the survey-derived temperature values for a thorough skill analysis of simulated surface and bottom temperature in the Bering10K model, an implementation of the Regional Ocean Modeling System (ROMS) covering the Bering Sea region. Overall, the Bering10K hindcast simulation captures observed patterns in eastern Bering Sea shelf bottom temperature well, with high correlation, low bias, and comparable interannual variability to the survey data. The exception to this is in the vicinity of the shelf break, where the model performed relatively poorly. This decrease in bottom temperature skill is attributable to bottom topography mismatches between the real and simulated shelf break location and is unavoidable in sigma-coordinate models like Bering10k; users should carefully consider the effects of the shelf break displacement whenever using model data extracted from this narrow region, particularly when attempting direct comparison with observations. Model skill was also generally higher in the southeastern portion of the shelf compared to the northern shelf region, though this may be an artifact of the low number of samples collected in the north relative to the south. Surface temperature performs with similar high correlation and comparable interannual variability, but simulations tended to be biased warm across much of the domain. This is likely due to a shallow bias in the simulated mixed layer that concentrates surface heating near the surface.

Contents

Introduction

As part of the annual assessment process for commercially important crab and groundfish species, the eastern Bering Sea shelf has been systematically surveyed via bottom trawl for the past four decades (Stevenson and Lauth 2012). Alongside the primary biological measurements (abundance, distribution, diets, and condition of groundfish and crab), a number of oceanographic measurements, including surface and bottom temperatures, are also collected during the trawls.

In recent years, a dataset derived from these survey observations of temperature has been used as a primary source of model validation for the temperature fields in the Bering10K ROMS model, a regional hydrodynamic model spanning the Bering Sea and northern Gulf of Alaska, with a focus on the eastern Bering Sea shelf (Kearney et al. 2020, Hermann et al. 2016). In addition, the sampling protocol used within the groundfish survey has served as a template for subsampling model simulations for use in a variety of research projects (e.g., Hollowed et al. 2020).

This report provides an overview of the data processing used to prepare the groundfish survey dataset for comparison with model results. It also provides a short discussion of the spatiotemporal variability that characterizes the survey observations, and the implications for model comparison and skill assessment. Finally, we provide an in-depth analysis of the skill performance of the Bering10K model relative to the observations from the bottom trawl survey dataset.

Groundfish Survey Data Analysis

Survey Sample Locations

Bottom trawl survey gear for the eastern Bering Sea shelf surveys was standardized in 1982, marking the start of the dataset considered in this study. The survey aims to resample the same locations each year at approximately the same time of year. However, the survey grid has not remained perfectly static over the entire 40-year survey period; instead, the full set of sampling stations has grown over the years to better quantify the primary species of interest. The original survey area, covering the southeastern shelf bounded by Unimak Pass to the south and St. Matthew to the north (Fig. 1, blue circles), encompasses the primary distributional area for groundfish and shellfish species. This region included 329 survey stations arranged on a 20-nautical mile (nmi) grid, and an additional 26 stations (Fig. 1, orange squares) at the corner points of the 20-nmi grid around St. Matthew and the Pribilof Islands for increased station density, designed to better sample blue king crab (*Paralithodes platypus*). In 1987, following high commercial landings of snow crab (*Chionoecetes opilio*) north of the existing survey region, the standard survey region was extended by adding 20 new stations to the northwest (Fig. 1, green diamonds). These three sets of stations encompass the current standard survey region, often referred to as the southeastern Bering Sea shelf (SEBS), and these stations have been systematically sampled via trawl every year through 2019. The northern portion of the shelf has been less consistently surveyed. Triennially between 1982 and 1991, trawls were conducted across parts of the northern shelf, with the goal of reassessing demersal fish and invertebrate stocks following an initial baseline survey in 1976 (Wolotira et al. 1977, Sample and Wolotira 1985, Goddard and Zimmerman 1993). The northern trawl stations were positioned on a 40-nmi grid between St. Matthew and St. Lawrence Islands (Fig. 1, purple +), plus a 10-nmi grid within Norton Sound

(Fig. 1, red dots). Between 1992 and 2009, sampling in the north was discontinued, though in 2005-2006, the northwestern samples did include the stations in stratum 81. In 2010, under the Alaska Fisheries Science Center's Loss of Sea Ice program (Sigler et al. 2015), sampling in the north was resumed, this time covering the entire northern shelf north to the Bering Strait and U.S.-Russia Maritime Boundary and using the same 20 nmi resolution as in the standard survey area (Fig. 1, gold x). The extended northern grid was discontinued in 2011 but resumed in 2017 and 2019, with plans to continue sampling biennially. Although sampling in the north region was not planned for 2018, the trawls collected in the standard survey area that year revealed very low numbers of walleye pollock (*Gadus chalcogrammus*) and Pacific cod (*G. macrocephalus*); this was similar to the patterns seen during the 2017 survey, when a northward latitudinal shift in these species was observed, likely due to historically warm conditions (Stevenson and Lauth 2019). Time constraints prevented a full survey of the northern region in 2018; instead, an additional 49 stations on an ad hoc 30-nmi grid were added to the 2018 survey (Fig. 1, pink stars).

The standard sampling plan design uses two vessels to conduct the trawl surveys. Sampling begins with vessels on adjoining columns in the eastern end of Bristol Bay, and both vessels sample alternate columns moving westward across the shelf (Fig. 2). This staggered sampling allows for calculation of the relative fishing power of the two vessels. The northern region is typically surveyed after the standard southeastern region is completed. The full survey takes approximately 2 to 4 months to complete each year. The majority of stations are sampled once per year, though a number of Bristol Bay stations are resampled toward the end of the survey to quantify molting and reproduction of red king crab (*Paralithodes camtschaticus*). Survey stations may also be resampled due to inadequacy of the biological measurements within a particular trawl; this can result in multiple temperature data points being collected at a particular station in a given year (Fig. 3).

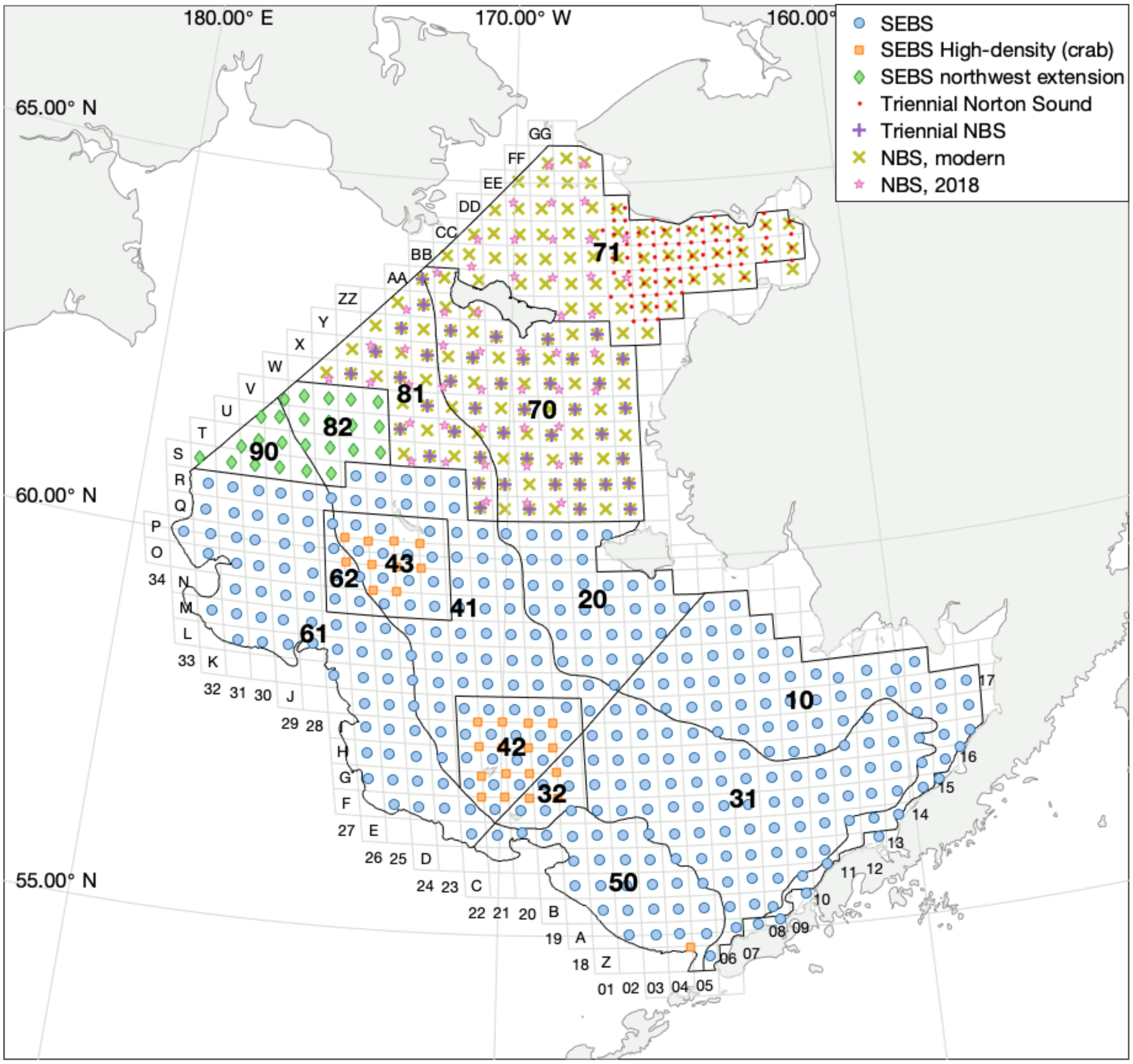

Figure 1. –– Map of Bering Sea groundfish survey sampling sites. The survey strata polygons (black lines, bold numbers) delineate biophysical regions for stratified sampling of the target groundfish species. The primary sampling sites lie on a 20-nautical mile grid (light gray), with each location identified by a station ID composed of row letter and column number. Colored markers indicate the mean sampling location for each station.

The equipment used to collect temperature data during these trawls has also evolved over time. From 1982 to 1989, temperature data were collected via expendable bathythermographs (XBTs). More recent surveys use digital bathythermograph recorders attached to the headrope of the bottom trawl net (BRANCKER RBR XL-200 Micro BTs recorded at 6-second intervals for the 1993-2001 surveys, and a Sea-Bird SBE-39 bathythermograph continuous data recorder at 3-second intervals for 2002 to present). Temperature is then averaged over the on-bottom and near-surface portions of the trawl (which covers a mean distance of 2.75 nmi) to produce a single surface and bottom (gear) temperature value per station per year. See Buckley et al. (2009) and Lauth et al. (2019) for full details of temperature data collection and post-processing.

Cleaning and Subsampling of Groundfish Survey data

We acquired survey data from two sources. The first was via a query of the RACEBASE database, the primary repository for the survey data; temperature data were queried via Oracle SQL, specifying all data after 1982 from the following Bering Sea surveys: Chukchi Sea Trawl Survey, Eastern Bering Sea Slope Bottom Trawl Survey, Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey, and Northern Bering Sea Crab/Groundfish Survey - Eastern Bering Sea Shelf Survey Extension (see Listing 1). The second data source was from a public archive of selected entries from RACEBASE, available in comma-delimited format via the AFSC website: https://apps-afsc.fisheries.noaa.gov/RACE/groundfish/survey_data/data.htm. Northern Bering Sea data collected prior to 2018 were only available to this author from the latter source.

We began our analysis by combining these two datasets. We identified duplicate sample points across the two data sources based on the cruise number, haul number, vessel number, and station ID; when a point was found in both datasets, the version from the RACEBASE query was kept and the public spreadsheet version was removed. Samples collected in the northern Bering Sea in 2018 were duplicated in the two datasets but labeled with different station IDs (in the RACEBASE query, these samples were labeled with non-standard station IDs indicating the use of the ad hoc 30 nmi grid, while in the public dataset they were auto-labeled with the nearest station from the standard 20 nmi grid); in this case, we also kept the RACEBASE version only. The public dataset points

Listing 1. – – RACEBASE query used to retrieve temperature data. The survey definition IDs correspond to the Eastern Bering Sea Crab/Groundfish Bottom Trawl Survey (98), Eastern Bering Sea Slope Bottom Trawl Survey (78), Northern Bering Sea Crab/Groundfish Survey - Eastern Bering Sea Shelf Survey Extension (143), and Chukchi Trawl Survey (6).

```
SELECT
    region,
    cruise,
    vessel,
    haul,
    to_char(start_time, ’YYYY-MM-DD␣HH24:MI:SS’)  start_time,
    surface_temperature  surface_temperature_celsius,
    gear_temperature     gear_temperature_celsius,
    abundance_haul
FROM
    racebase.haul
WHERE
    cruisejoin IN (
        SELECT
            racebase_cruisejoin
        FROM
            race_data.cruises
        WHERE
            survey_id IN (
                SELECT
                    survey_id
                FROM
                    race_data.surveys
                WHERE
                    survey_definition_id IN ( 98, 78, 143, 6 )
                    AND year >= 1982
```

were geolocated with a single latitude and longitude coordinate per sample, while the RACEBASE version included coordinates for both the start and end trawl position. We calculated a mean latitude and longitude value for the RACEBASE values by averaging these two positions with a simple non-geographic mean. We treated the START_TIME and DATETIME fields in the RACEBASE and public datasets, respectively, as identical fields, and removed the redundant YEAR field from the public dataset. We also renamed the SURF_TEMP and BOT_TEMP fields in the public dataset as SURFACE_TEMPERATURE and GEAR_ TEMPERATURE, respectively, to match the RACEBACE data. The public data did not include the HAUL_TYPE field, which would be used in our later analysis, so we marked these samples with a novel value (24). Any other missing fields were left empty. Finally, any entries that did not include data for either the GEAR_TEMPERATURE or SURFACE_TEMPERATURE fields were removed from the combined dataset.

From this combined dataset, a few cruises were removed. The Chukchi Sea is outside the domain of the Bering10K ROMS model, so all cruises from this region were removed. Within the Bering10K ROMS model, the continental slope rises less steeply than in reality. This is due to bathymetric smoothing that is necessary to avoid numerical issues in the model, but as a result, direct comparison of modeled bottom temperature to observations becomes more complicated in the region of the continental slope since the mismatch in depth between the real world and modeled world can be several hundred meters. Therefore, all data collected from the slope survey (survey ID 78) were also removed (data from near the shelf break and slope but collected during the shelf survey remained; we address these points later). Finally, in February 1983, a short survey was conducted with samples collected in Bristol Bay, near the southern shelf break, and near the Pribilof Islands. This is the only instance of data being collected so early in the year; the dataset is too small to provide any model skill assessment, so this cruise (cruise 198301, vessel 21) was also removed from the final temperature dataset.

As mentioned earlier, the northern Bering Sea survey data from the 80s and 90s were available only in the public dataset. Like the 2018 northern data from the public dataset, these points had been labeled with auto-assigned station IDs from the 20-nmi grid, with each sample assigned to the station closest to its collection point coordinates. However, the samples collected within Norton Sound were actually collected based on a 10-nmi grid. We were unable to locate any description of this grid and the coordinates of the target sample locations associated with it beyond the hand-drawn maps in Wolotira et al. (1977). Instead, we reconstructed the grid using the trawl sample coordinates. We applied a k-medioids clustering to the sample coordinates, identifying groups of sample points across years that were located near each other. This method identified 81 unique locations that fell approximately on a 10-nmi grid (with wider spacing farther inshore); the number of clusters was chosen by trial and error, and there remained some ambiguity among the more scattered points on the western edge of the Norton Sound region, but we considered this labeling sufficient for our purposes. These stations were assigned station IDs with the format of `norton01`, `norton02`, ... `norton81`, with the numeric portion assigned at random (but with a prescribed random seed for repeatability). These Norton Sound stations were all labeled as part of stratum 71.

A number of small adjustments were made to the station ID field of certain entries. In general, stations located on the 20-nmi grid use a format of <row>-<col>, where <row> consists of a single alphabetic character, or two repeating alphabetic characters corresponding to the grid row, and <col> is a two-digit number corresponding to the grid column (see Fig. 1). A few database entries included variants, such as using single-digit numbers or preceding the row character with one or more 0s; these were adjusted to match the expected format. Corner-of-grid stations followed a naming convention of <row1><row2><col1><col2> based on the adjacent columns and rows; a handful of stations listed these with a hyphen in the station ID, and these were also standardized to the more common no-hyphen format.

A few of the entries were missing data in the `STRATUM` and `STATION_ID` fields. When a sample included a station ID but no stratum number, we assigned that point the same stratum value as other points with that same station ID. A few stations along the U.S./Russia border were always listed without a stratum value. For our analysis, stations S-32, T-31, and U-30 were assigned to stratum 90, and station V-29 was assigned to stratum 82. A small handful of remaining entries did not include any station ID and were assigned one based on proximity to the mean sampling location of each station; these stations were also assigned stratum values as needed based on shared station IDs. Once all entries were assigned a station ID and stratum value, the dataset was reviewed to check for inconsistent values. Data from station W-22 were sometimes labeled as in stratum 70 and sometimes in stratum 81; more recent years favored the 81 designation so we opted for the same convention. A handful of 2018 northern stations were marked as stratum 70 while actually falling in 71 and were corrected. Any entry whose STRATUM or STATION_ID value was changed from the original dataset was marked as such in the FLAGSTRATUM and FLAGSTATION fields, respectively.

For many of our analyses, we wanted to include only a single sample point from each station in each year. To simplify those calculations, a final column was added to the dataset (BESTREP) indicating which samples were considered the best representative for each station/year combination. For this, a sample was preferred if it was marked with a HAUL_TYPE of 3, indicating a standard bottom sample at a preprogrammed station, and if the sample performance was marked as good (0). If no samples meeting this criteria were found, preference was given to good performance samples with any haul type, followed by satisfactory performance samples, and finally unsatisfactory performance samples (under the assumption that temperature data remained valid despite unsatisfactory performance of the tow for groundfish sampling purposes).

See Table 1 for a full description of the variables in the final dataset. This dataset is available in the accompanying AFSC_groundfish_survey_temperature_1982-2020.xlsx spreadsheet (DOI: 10.5281/zenodo.4567557) under the SurveyData sheet.

A second table was constructed holding summary information related to each sample station. This table included all stations that fell on either the primary 20-nmi grid, the 10-nmi Norton Sound grid, or the 30-nmi northern Bering Sea grid. Stations that were sampled for other purposes were not included in the summary table though they remain in the primary dataset. Also, stations A-01 and D-11 were removed from the summary table; these stations are both located near the edge of the sampling region, and have been sampled only sporadically. This summary table is available in the accompanying AFSC_groundfish_survey_temperature_1982-2020.xlsx spreadsheet under the StationSummary sheet.

Summary of Groundfish Survey-derived Datasets

Table 1. –– Survey variable descriptions. The majority of the variables and descriptions reflect the RACE database codes as documented in the RACE codebooks (RACE 2019a,b). Primary variables used for most model-to-data comparisons are indicated with an asterisk.

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td>REGION</td><td>Region. In this dataset, all are BS for Bering Sea.</td></tr>
<tr><td>VESSEL</td><td>Vessel code indicating ship used for trawl. In this dataset, codes are as follows:
<table>
<tr><th>Code</th><th>Vessel</th></tr>
<tr><td>1</td><td>RV <i>Chapman</i></td></tr>
<tr><td>19</td><td>MV <i>Pat San Marie</i></td></tr>
<tr><td>21</td><td>RV <i>Miller Freeman</i></td></tr>
<tr><td>37</td><td>RV <i>Alaska</i></td></tr>
<tr><td>57</td><td><i>Morning Star</i></td></tr>
<tr><td>60</td><td><i>Argosy</i></td></tr>
<tr><td>78</td><td><i>Ocean Hope 3</i></td></tr>
<tr><td>87</td><td><i>Tracy Anne</i></td></tr>
<tr><td>88</td><td>FV <i>Arcturus</i></td></tr>
<tr><td>89</td><td>FV <i>Aldebaran</i></td></tr>
<tr><td>94</td><td>FV <i>Vesteraalen</i></td></tr>
<tr><td>134</td><td><i>Northwest Explorer</i></td></tr>
<tr><td>162</td><td><i>Alaska Knight</i></td></tr>
</table></td></tr>
<tr><td>CRUISE</td><td>Cruise number, with format YYYYNN, where YYYY is the year and NN is the index indicating order for that year</td></tr>
<tr><td>HAUL</td><td>Haul number</td></tr>
<tr><td>HAUL_TYPE</td><td>Haul type. Within this dataset, the following codes are used</td></tr>
</table>

Table 1. –– Survey variable descriptions (continued.)

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td></td><td>
<table>
<tr><th>Code</th><th>Description</th></tr>
<tr><td>0</td><td>opportunistic (not a programmed station)</td></tr>
<tr><td>3</td><td>standard bottom sample</td></tr>
<tr><td>4</td><td>fishing power comparative sample</td></tr>
<tr><td>5</td><td>commercial prospect sample</td></tr>
<tr><td>6</td><td>trawl on predetermined trackline targeted on fish as encountered</td></tr>
<tr><td>7</td><td>fishing gear experiment (not quantitative)</td></tr>
<tr><td>8</td><td>opportunistic off-bottom sample</td></tr>
<tr><td>9</td><td>tow for tag and release</td></tr>
<tr><td>13</td><td>index sample tow</td></tr>
<tr><td>15</td><td>unknown</td></tr>
<tr><td>17</td><td>crab resurvey tow</td></tr>
<tr><td>18</td><td>crab experimental tow</td></tr>
<tr><td>19</td><td>crab hot spot tow</td></tr>
<tr><td>20</td><td>catch selective sampled/processed</td></tr>
<tr><td>21</td><td>yellowfin sole nearshore station</td></tr>
<tr><td>24</td><td>.csv import (added for this report, not an official RACE code)</td></tr>
</table>
</td></tr>
<tr><td>PERFORMANCE</td><td>Trawl performance codes, where 0 indicates good performance, positive codes indicate satisfactory performance, and negative codes indicate unsatisfactory performance. See Appendix i in RACE (2019a) for subcategories.</td></tr>
<tr><td>START_TIME</td><td>Trawl start time, formatted as a DD-MMM-YY character array</td></tr>
<tr><td>DURATION</td><td>Trawl duration, in hours</td></tr>
<tr><td>DISTANCE_FISHED</td><td>Distance fished, in km</td></tr>
<tr><td>NET_WIDTH</td><td>Width of net opening, in m</td></tr>
<tr><td>NET_MEASURED</td><td>Flag indicating whether net was measured ([Y]es/[N]o)</td></tr>
<tr><td>NET_HEIGHT</td><td>Height of net opening, in m</td></tr>
<tr><td>*STRATUM</td><td>Survey stratum where sample is located (see Fig. 1)</td></tr>
<tr><td>START_LATITUDE</td><td>Latitude at start of trawl</td></tr>
<tr><td>END_LATITUDE</td><td>Latitude at end of trawl</td></tr>
<tr><td>START_LONGITUDE</td><td>Longitude at start of trawl</td></tr>
<tr><td>END_LONGITUDE</td><td>Longitude at end of trawl</td></tr>
<tr><td>*STATIONID</td><td>Station ID (see Fig. 1)</td></tr>
<tr><td>GEAR_DEPTH</td><td>Depth of gear during trawl, in m</td></tr>
<tr><td>BOTTOM_DEPTH</td><td>Bottom depth, in m</td></tr>
<tr><td>BOTTOM_TYPE</td><td>Bottom substrate type (Note: two-digit codes are no longer actively used but are still attached to some older samples)</td></tr>
</table>

Table 1. –– Survey variable descriptions (continued.)

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td></td><td>
<table>
<tr><th>Code</th><th>Description</th></tr>
<tr><td>0</td><td>Unidentified</td></tr>
<tr><td>1</td><td>Mud</td></tr>
<tr><td>2</td><td>Clay</td></tr>
<tr><td>3</td><td>Sand</td></tr>
<tr><td>4</td><td>Gravel</td></tr>
<tr><td>5</td><td>Cobbles</td></tr>
<tr><td>10</td><td>Grey mud</td></tr>
<tr><td>11</td><td>Grey clay</td></tr>
<tr><td>12</td><td>Mud and clay</td></tr>
<tr><td>30</td><td>Green mud and sand</td></tr>
<tr><td>31</td><td>Mud and sand</td></tr>
<tr><td>49</td><td>Grey sand and worm tubes</td></tr>
<tr><td>51</td><td>Sandy</td></tr>
<tr><td>52</td><td>Grey sand</td></tr>
<tr><td>54</td><td>Black sand</td></tr>
<tr><td>55</td><td>Grey sand, mud, gravel</td></tr>
<tr><td>59</td><td>Gravel and sand</td></tr>
<tr><td>62</td><td>Rocky</td></tr>
<tr><td>63</td><td>Gravel</td></tr>
<tr><td>72</td><td>Hard clay</td></tr>
<tr><td>74</td><td>Hard</td></tr>
</table>
</td></tr>
<tr><td>*SURFACE_TEMPERATURE</td><td>Temperature at surface, in °C</td></tr>
<tr><td>*GEAR_TEMPERATURE</td><td>Temperature at depth of trawl gear, °C</td></tr>
<tr><td>WIRE_LENGTH</td><td>Length of wire, in m</td></tr>
<tr><td>GEAR</td><td>Trawl gear type. See Gear Code Table in RACE (2019b) for further details:</td></tr>
</table>

Table 1. –– Survey variable descriptions (continued.)

Variable	Description	
	Code	Description
	20	400-Mesh eastern trawl with 94’ footrope and 71’ headrope, path width is 12.19 m
	26	Same as 20, but path width = 47’
	30	Eastern trawl with 112’ footrope and 83’ headrope
	33	Same as 30, but path width = 54.64’
	34	Same as 30, but path width = 53.36’
	35	Same as 30, but path width = 59.00’
	37	Same as 30, but path width = 54.264’
	38	Same as 30, but path width = 53.852’
	39	Same as 30, but path width = 59.055’
	40	Same as 30, but path width = 54.068’ (16.48 m) and vertical opening = 3.0 m.
	42	Same as 30, but path width = 54.71’ (16.67 m) in depths less than 100 m.
	43	Same as 30, but path width = 58.41’ (17.80 m) in depths greater than 100 m.
	44	Same as 30. Acoustic net mensuration equipment attached.
	45	400 Mesh Alaska Department of Fish & Game (ADF&G) Eastern trawl. 78’ headrope 95’ footrope.
	160	Nor’eastern trawl, 90’ headrope, 105’ footrope.
	172	Poly-nor’eastern, four seam, hard bottom, high rise rockfish trawl constructed of polyethylene.
	219	3-m beam trawl
ACCESSORIES		

Table 1. –– Survey variable descriptions (continued.)

Variable	Description	
	Code	Description
	2	6’×9’ steel v-doors, 25 fm dandylines branching to 15 fm bridle. 1.25” Codend liner, no chains.
	15	6’×9’ steel v-doors (standardized to 1,800 lbs after 1988), double 30 fm 5/8” dandylines, 1.28” Mesh codend liner, 24” chain extension between lower dandyline and footrope.
	34	5’×7’ steel v-doors, 25 fm dandylines (15 fm single, 10fm double), 18”×8” floats on headrope, 1.25” Mesh liner in codend, no weight on footrope.
	47	6’×9’ steel v-doors, 40 fm dandylines (25 fm single, 15 fm double), 1.25” Liner, no roller gear.
	57	6’×9’ steel v-doors, 2,200lbs each. Three 30 fm, 5/8” galvanized bridles from each side. West coast slope survey modified roller gear (8” diameter solid rubber disks, strung from wing to wing on 5/8” high tensile chain for added weight) and 1/2” long link chain fishing line.
	64	Net rigging consists of triple 180’ (54.9 m), 5/8” (1.6 cm) diameter galvanized wire rope dandylines. Dandylines are rigged with 18” and 9” chain extensions to the headrope and side panel attachments respectively. Steel v-doors, 6’×9’ (1.83×2.74 m), weighing from 1,300 to 2,200 lbs each are standard. The roller gear is 79’ 6” (24.2 m) long and constructed of 3/4” (1.91 cm) 6×9 galvanized wire rope, 14” (36 cm) rubber bobbins separated by a solid string of 4” (10 cm) rubber disks. In addition, 19’ 6” (5.9 m) wire rope extensions with 4” (10 cm) and 8” (20 cm) rubber disks were used to span each lower flying wing section. Polypropylene chafing gear: 10” (25.3 cm) mesh of 3/8” (1 cm) poly rope hog ringed or interwoven, 46 mesh circum. By 21.5 Mesh deep, laced to outer bag.
	140	Beam trawl. 3” Pipe frame with semicircle 3” flat strap end runners. 7 ft wide overall × 2 ft high. 1 1/4 inch nylon net with 118 inch footrope. 5/16 proof coil chain weight sewn on footrope. Net 22 ft overall with 1/2 inch knotless cod end. (Used on Arcturus cruise 199801 towed behind 83/112 trawl with an underbag)
SUBSAMPLE	Subsampling method	

Table 1. –– Survey variable descriptions (continued.)

<table>
<tr><th>Variable</th><th>Description</th></tr>
<tr><td></td><td><table>
<tr><th>Code</th><th>Description</th></tr>
<tr><td>0</td><td>Catch not processed</td></tr>
<tr><td>1</td><td>No subsampling</td></tr>
<tr><td>2</td><td>Catch Subsampled - Load Cell</td></tr>
<tr><td>3</td><td>Catch Subsampled - Volumetric</td></tr>
<tr><td>4</td><td>Catch Subsampled - Visual Estimate</td></tr>
<tr><td>5</td><td>Unknown</td></tr>
<tr><td>6</td><td>Catch Subsampled - Basket Weight</td></tr>
<tr><td>7</td><td>Not recorded</td></tr>
<tr><td>9</td><td>Non-quantitative Catch Sampling</td></tr>
<tr><td>11</td><td>Selective catch sampling for quantitative purposes</td></tr>
<tr><td>12</td><td>Catch subsampled, without load cell weight of catch. Subsample fraction was estimated by volumetric method. Density Lookup Table 2014 was used.</td></tr>
<tr><td>13</td><td>Catch subsampled, without load cell weight of catch. Subsample fraction estimated by volumetric method. Density was calculated on deck from haul sample.</td></tr>
</table></td></tr>
<tr><td>ABUNDANCE_HAUL</td><td>Flag indicating whether adundance was measured ([Y]es/[N]o)</td></tr>
<tr><td>*LATITUDE</td><td>Latitude, mean of start and end location</td></tr>
<tr><td>*LONGITUDE</td><td>Longitude, mean of start and end location</td></tr>
<tr><td>*DATETIME</td><td>Excel serial date number corresponding to start time</td></tr>
<tr><td>FLAGSTATION</td><td>Flag indicating whether station ID has been modified from the original (true/false)</td></tr>
<tr><td>FLAGSTRATUM</td><td>Flag indicating whether stratum number has been modified from the original (true/false)</td></tr>
<tr><td>TYPE</td><td>Station type, based on text-parsing of the station ID<table>
<tr><th>Code</th><th>Description</th></tr>
<tr><td>0</td><td>other (not one of the below)</td></tr>
<tr><td>1</td><td>main grid, row letter-column number combination (e.g., A-01)</td></tr>
<tr><td>2</td><td>northern Bering shelf, NBS-X (e.g., NBS-1)</td></tr>
<tr><td>3</td><td>corner, two row letters and two column numbers (e.g., AZ0504)</td></tr>
<tr><td>5</td><td>samples marked as being collected to the north, south, east, or west of an existing station (e.g, G-21E)</td></tr>
<tr><td>6</td><td>norton, from the Norton sound 10-n mi grid, with IDs created specifically for this study (e.g., norton01)</td></tr>
<tr><td>8</td><td>IDs starting with SP-XX, indicating special projects and typically associated with yellowfin sole sampling (e.g., SP-01)</td></tr>
</table></td></tr>
<tr><td>*BESTREP</td><td>Flag indicating whether sample is the best representative for its year and station</td></tr>
</table>

Table 2. – – Descriptions of variables found in the StationSummary table.

Variable	Description
STATIONID	Station ID
LATITUDE	Mean sampling latitude over the 1982-2019 surveys
LONGITUDE	Mean sampling longitude over the 1982-2019 surveys
TYPE	Grid on which station is located, either main (20 nmi grid), corner (of 20 nmi grid), nbs (30 nmi grid), or norton (10 nmi grid))
STRATUM	Stratum in which station is located
DOY	Mean day of year when station was sampled over the 1982-2019 surveys
B10K_XI	ξ-axis coordinate of Bering10K ROMS domain grid cell located closest to the mean sampling location
B10K_ETA	η-axis coordinate of Bering10K ROMS domain grid cell located closest to the mean sampling location

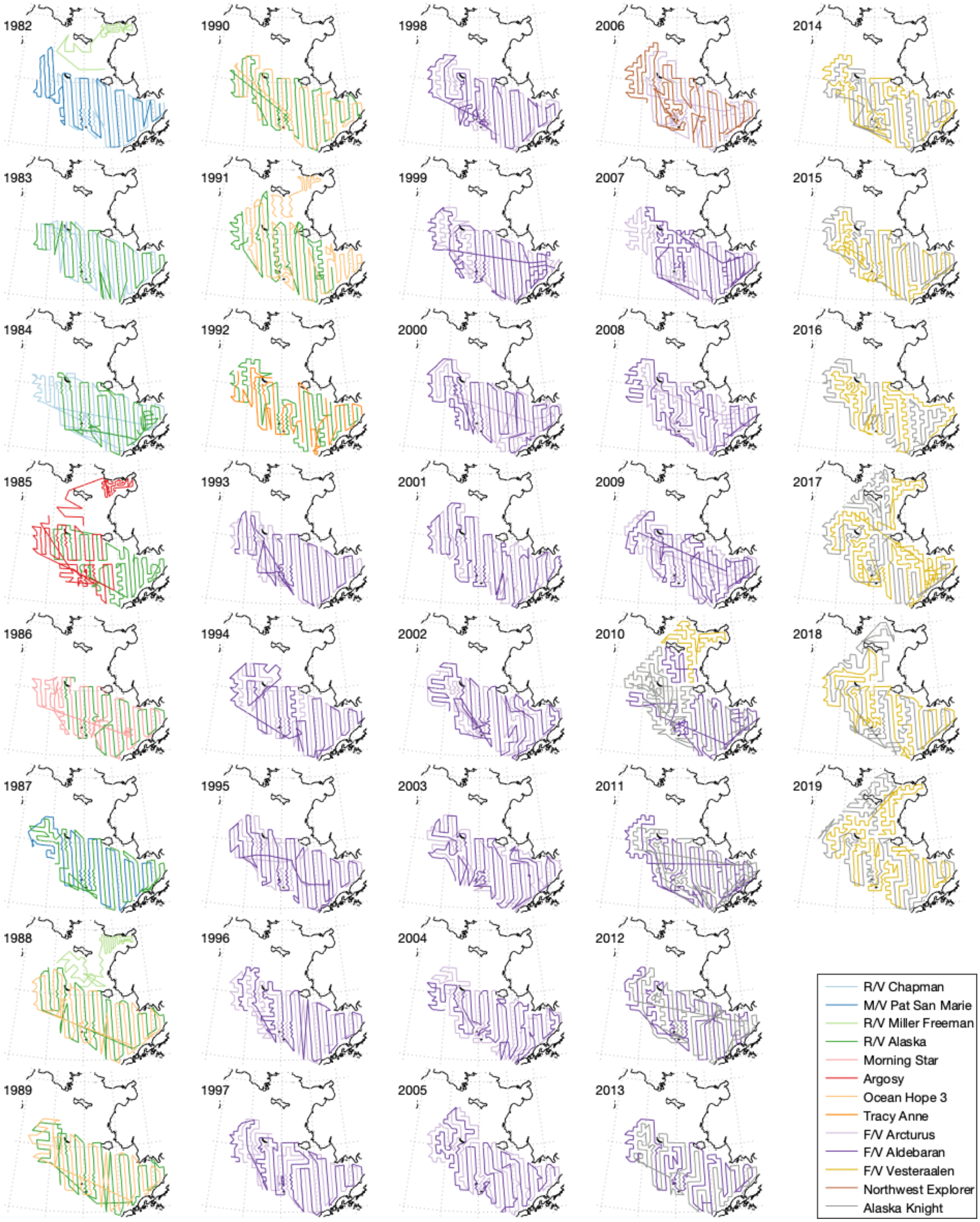

Figure 2. –– Survey cruise tracks by year, colored by survey vessel.

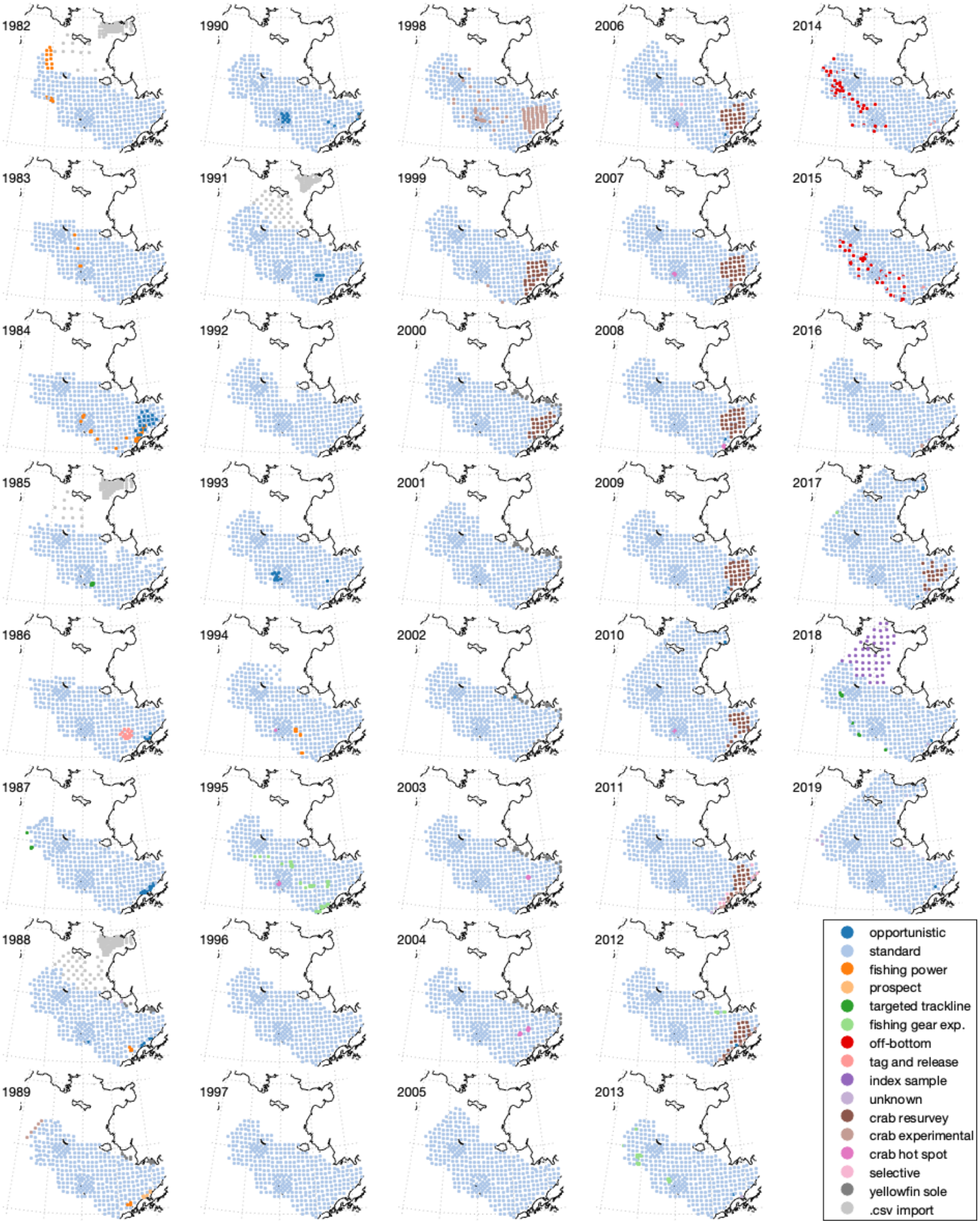

Figure 3. –– Survey sampling sites by year, categorized by haul type. See Table 1 for more information on these various haul types.

Resampling a Model for Comparison With Survey Dataset

Survey Replication Methods

We use two different methods for extracting data from a model for comparison with this groundfish survey-derived temperature dataset. For the purposes of this document, we will describe the specifics applicable to the Bering10K ROMS model, though the methods could be easily adapted to any similar hydrodynamic model.

The first method, often referred to as survey replication, involves extracting an analogue sample corresponding to each point in the SurveyData dataset. Each SurveyData sample is matched up to the model grid cell whose rho-coordinates (i.e., coordinates of the center of the grid cell) are closest to the sample trawl mean latitude and longitude. Likewise, each trawl sample is matched to the model output time slice closest to the trawl date and time. By matching each individual point, this resampling aims to replicate the same spatial and temporal variability as seen in the groundfish survey. This direct matching across space and time can only be applied to a simulation that covers the same period as the trawl survey itself (i.e., 1982 to 2019); it is most useful when applied to a hindcast simulation that is designed to capture realistic interannual variability. A survey-replicated dataset derived from the Bering10K hindcast simulation is used for the skill assessment described in the next section of this report.

The second method of model sampling is sometimes referred to as idealized or climatological survey replication. This method is similar to the first one, but it is based on the StationSummary data rather than the raw survey sample data. From each year of a simulation, points are extracted from the grid point closest to the mean station location and on the mean day of year that station was sampled. Because this method uses day of year rather than specific dates, it can be applied to simulations that span any time period, including future projections. The objective of this sampling method is to create a dataset that captures the spatiotemporal characteristics of the groundfish survey, even outside of the specific years of survey. It can be particularly useful when combining model simulations with empirical models that are based on the biophysical relationships derived from the groundfish survey data (e.g., Holsman et al. (2020)).

Spatiotemporal Variability in the Groundfish Survey Data

Over the summer period when the survey is conducted, the middle and outer shelf regions (Fig. 1, strata 31–62 and 81–90) are strongly thermally stratified. The bottom temperatures in this region remain relatively constant over the entire survey period because the deeper waters are isolated from surface heating (Fig. 4). These strata also tend to be surveyed over a relatively narrow range of time each year. As a result, the year-to-year differences seen in observed temperatures in these particular regions can be attributed primarily to interannual variability stemming from variations in seasonal ice cover. In contrast, the shallower inner shelf regions (Fig. 1, strata 10, 20, 70, and 71) are well-mixed throughout the water column, with little to no stratification in the summer. Consequently, the bottom temperatures in this region experience seasonal warming as cold, ice-influenced waters are warmed by surface heating (Fig. 4). Therefore, the variations in bottom temperature seen in these regions are a function both of temperature variability between years and variability between the time of year in which samples were collected. Within this dataset, interannual variability tends to be higher than the within-year variability, but the latter can still account for a large portion of the overall variability.

The seasonal variations in temperature, especially along the inner shelf, are also visible in the composite temperature maps that are often used to display survey-measured bottom temperatures. For example, bottom temperatures from the 2010 survey (Fig. 5, panel a) show a clear north/south gradient in temperature along the inner domain, with warmer temperatures in the north. The Bering10K model, when subsampled at the same times of year as the 2010 trawl (Fig. 5, panel b) shows a very similar pattern. However, constant-time slices of bottom temperature extracted from the model on the first and last days of the 2010 survey, i.e., 7 June and 10 Aug 2010 (Fig. 5, panels c and d, respectively), indicate that this gradient is entirely an artifact of the time of year when the northern stations were sampled relative to the southern ones. In fact, the southern regions appear to be warmer than the northern ones at both time snapshots. When comparing spatial patterns in the Bering10K-simulated and trawl-derived bottom temperatures, one needs to be careful to consider the within-year variations present in the trawl data. This variability should also be kept in mind when comparing any idealized survey-replicated model data to the actual survey data (Fig. 6).

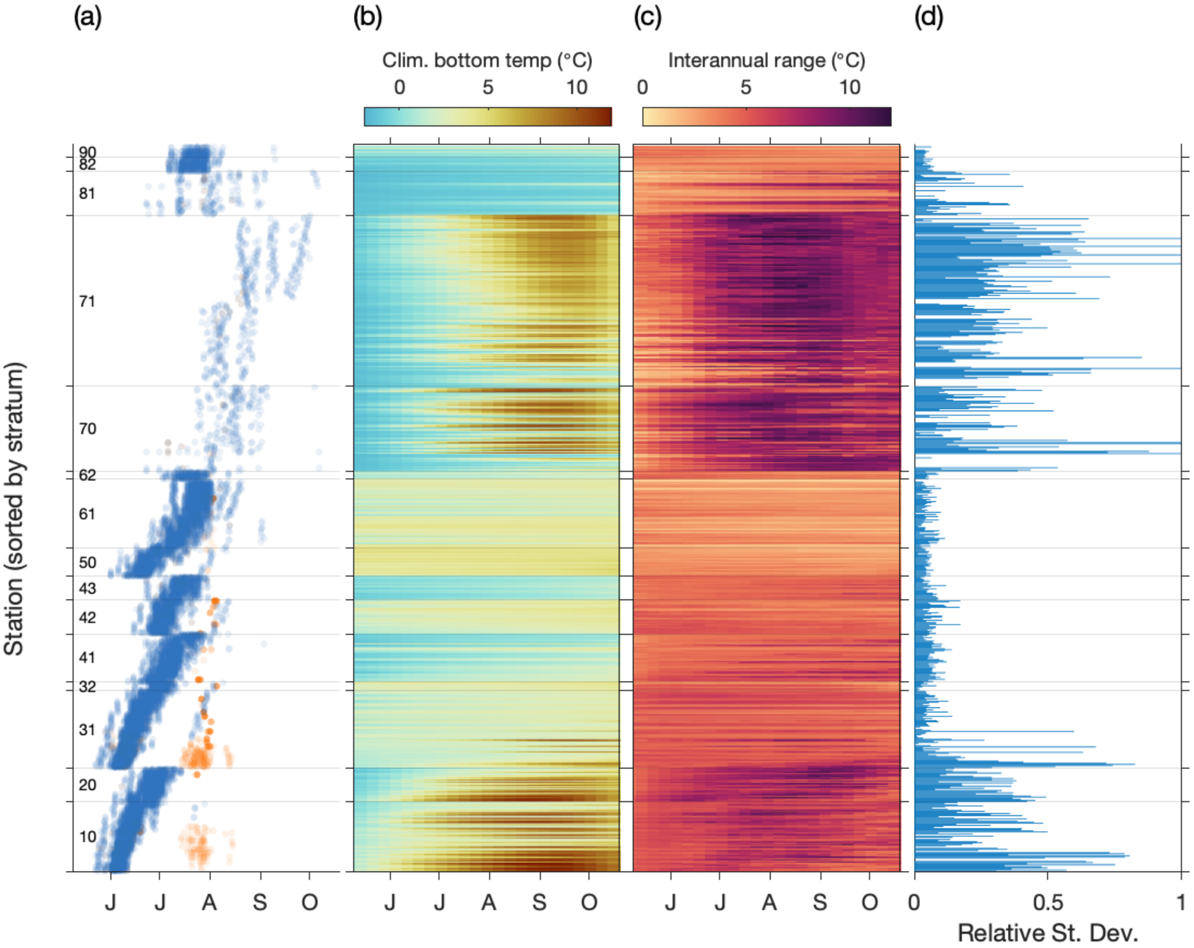

Figure 4. –– Variations in bottom temperature relative to sampling date across the Bering Sea survey region. a) Dots indicate each individual trawl sample, organized by station on the y-axis (sorted by the stratum in which each station is located) and time of year collected on the x-axis. Blue dots indicate the primary trawl at each location, and orange dots indicate additional trawls (incomplete trawls, replications, etc.). b) Climatological (Clim.) hindcast bottom temperature versus time of year at each station location. c) Interannual range in bottom temperature versus time of year at each station location. d) Fraction of observed variability that could be due to sampling time variations, defined as the standard deviation of survey-replicated bottom temperatures from the Bering10K hindcast divided by the standard deviation of the same sampling applied to a climatological bottom temperature time series.

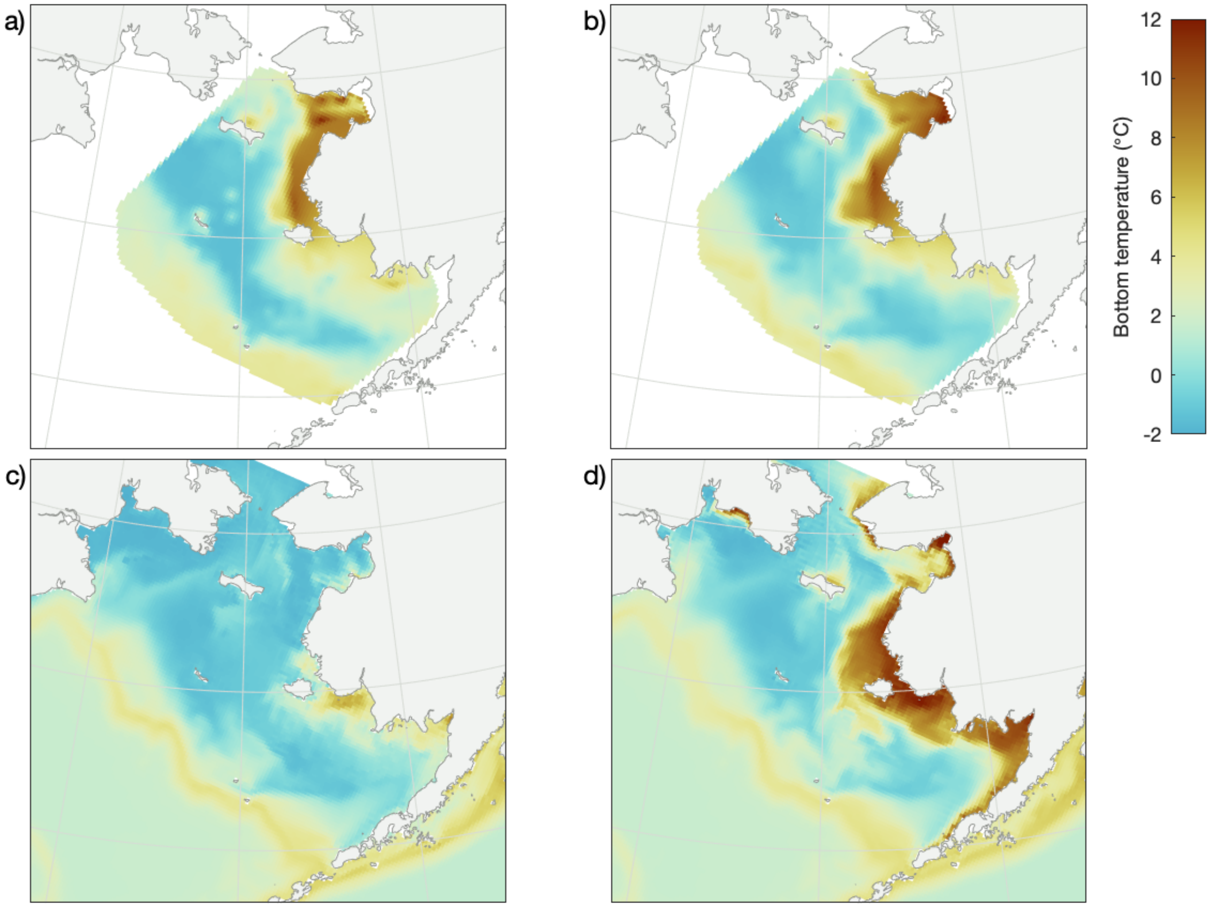

Figure 5. −− a) A composite of bottom temperature based on trawl-sampled survey points from 2010, interpolated to the model grid via a natural neighbor interpolant. b) A composite of bottom temperature based on the Bering10K hindcast sampled using the survey replication method, interpolated to the model grid using the same interpolant as in (a). c) Bering10K bottom temperature on 7 June 2010 (first day of the 2010 survey). d) Bering10K bottom temperature on 10 Aug 2010 (last day of the 2010 survey).

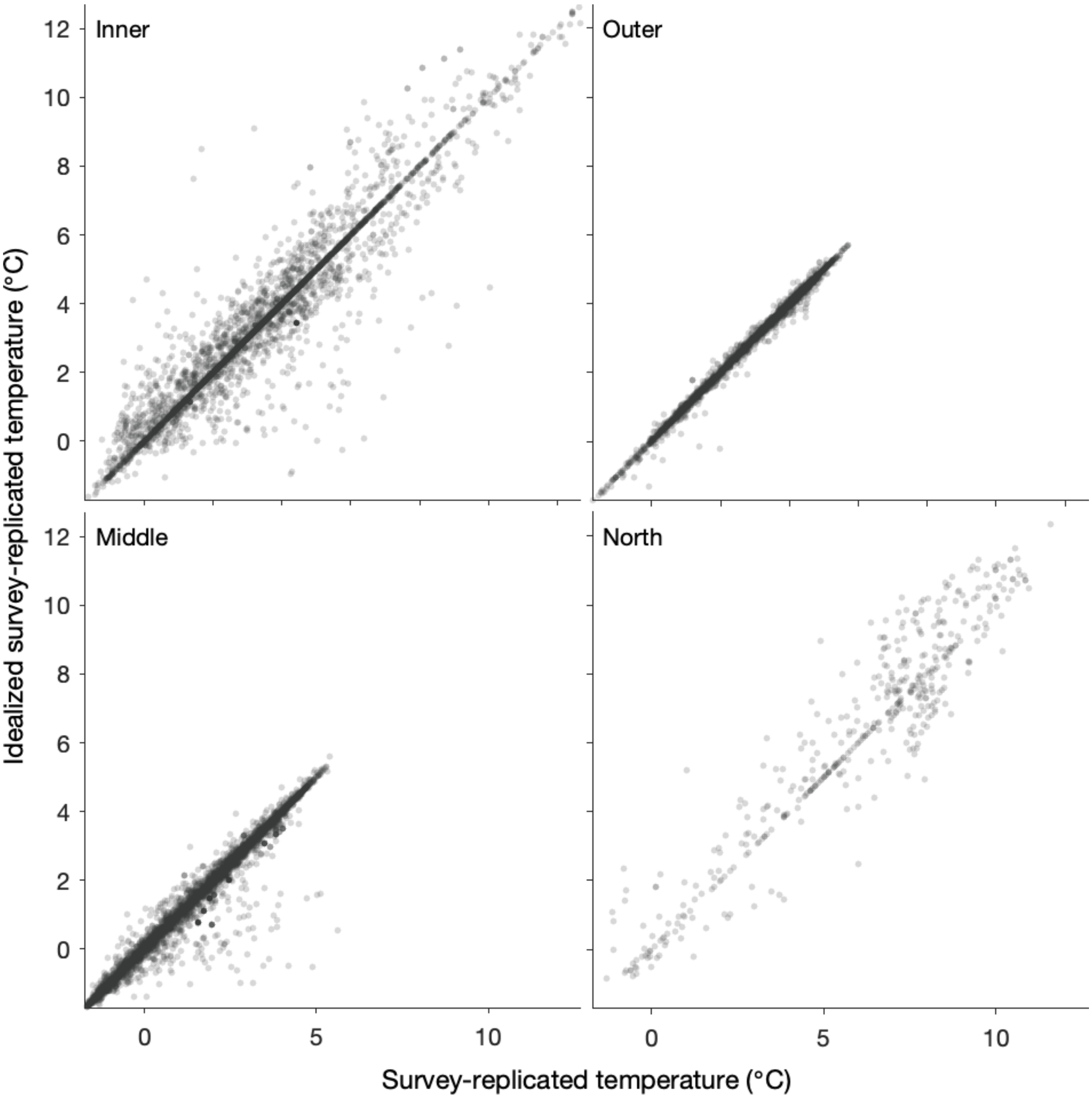

Figure 6. –– A comparsion of survey-replicated bottom temperature extracted from the Bering10K hindcast simulation, compared with an idealized survey-replication of the same simulation, across different regions of the shelf domain (see Fig. 7 for a map of the four regions; here we combine the outer and shelf break regions together).

Bering10K ROMS Skill Assessment of Temperature

The Bering10K model is an implementation of the Regional Ocean Modeling System, a free-surface, primitive equation hydrographic model (Shchepetkin and McWilliams 2005, Haidvogel et al. 2008). The Bering10K ROMS domain spans the Bering Sea and northern Gulf of Alaska, with 10 km horizontal resolution; it was first developed as part of the Bering Ecosystem Study (BEST) and Bering Sea Integrated Ecosystem Research Project (BSIERP), and it has since been used in a number of studies (Hermann et al. 2013, Kearney et al. 2020, and citations within).

Kearney et al. (2020) presents a thorough validation of the Bering10K ROMS model, focusing on the physical and biogeochemical variables of interest in many of the ecosystem modeling studies where Bering10K is currently being used. Within the Kearney et al. (2020) study, skill related to bottom temperature, including assessments of the simulated cold pool, were quantified relative to the groundfish survey temperature dataset described in this report. An assessment of bottom temperature seasonal forecast skill (Kearney et al., in prep.) likewise presents similar skill metrics to support its use of the a hindcast simulation. Here, we provide a more in-depth look at the calculations underlying those skill assessments. We refer readers to the aforementioned publications for greater context of the use of these skill assessments; here we focus on more detailed specifics.

The Bering10K hindcast simulations that we evaluated here were driven by surface atmospheric forcing from a collection of reanalysis products: the Common Ocean Reference Experiment (CORE; Large and Yeager 2009), the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), and the Climate Forecast System Operational Analysis (CFSv2). The CORE dataset included input forcing from 1970 to 2003. CFSR spans 1979 to March 2011. The CFSv2 Operational Analysis data began in April 2011 and continued as an operational product to the present. The first hindcast simulation in this study, which was the primary one used for research simulations, uses a combination of these datasets in order to span the longest possible period. The model used CORE input from 1970 to 1995, then switched to CFSR. To account for small mismatches in downwelling radiation between the two products, the CORE shortwave and longwave radiation values were divided by factors of 0.9 and 0.97, respectively. We also ran simulations using just the CORE forcing (1970 to 2003) and just the CFS forcing (1979 to 2018), without any adjustments to radiation values. The three simulations performed comparably, so we excluded further in-depth analysis of these three variants (but separate skill statistics for each are provided for reference.)

Hindcast skill was assessed for model bottom and surface temperature versus the groundfish survey gear temperature and surface temperature, respectively. Model bottom temperature was defined as the mean temperature over the bottom 5 m of the model domain, while model surface temperature was defined as the mean temperature of the top 5 m relative to the free surface. Skill statistics were first calculated by station for all stations that had been sampled in at least three different years since 1982 (Fig. 8 and Fig. 9). The stations were then divided into five biophysical regions: inner shelf (strata 10–20,70), middle shelf (strata 31–43,81–82), northern/Norton Sound (strata 71), shelf break (stations beyond the modeled 200-m contour), and outer shelf (strata 50–62, 90 except shelf break stations) (Fig. 7). A sixth region, encompassing the primary southeastern Bering Sea survey region (strata 10–62) was added to these five; this final region is the one typically used for calculation of the Bering 10K-derived cold pool index. For all six regions, skill statistics were first calculated on a point-by-point basis, comparing all survey station observations within each region across time to their survey-replicated modeled counterparts; these statistics are presented for each station and across each region (i.e., regionally grouped). A separate calculation was performed using regional averages; for this, we eliminated any within-year repeat samples (i.e., used only those labeled as best-replicates as described in section) and then averaged spatially to create annual time series for both the survey and survey-replicated datasets (Figs. 10 and 11).

Overall, the hindcast simulation bottom temperature captures the patterns seen in each biophysical region well, with high correlation, low bias, and comparable interannual variability to the survey data (Tables 3 and 4). The exception to this is the shelf break region, which performs relatively poorly. In this region, the model requires bathymetric smoothing to avoid errors in the horizontal pressure gradient that are characteristic of sigma-coordinate models like ROMS in areas of steep topography. Because of this, the modeled shelf is slightly narrower than the real world one, and the survey-replicated locations that fall between the simulated and real shelf break end up comparing the simulated slope to the real world shelf; we do not expect these data points to be directly comparable, and caution against using these locations in any model-to-data comparison. In general, those using Bering10K model output from the vicinity of the shelf break and slope should remain aware of the displacement of the shelf break and adjust their calculations accordingly.

The station by station skill analysis revealed some variation in the model's ability to capture spatiotemporal patterns in bottom temperature. A cool bias near the 50-m contour suggests that the model may place the inner front farther inshore than was seen in observations. Skill was also generally lower in the northern regions, though this may be an artifact of the low number of samples collected there.

The simulated surface temperature also performed well across regions in terms of capturing interannual variability. However, the simulated surface temperature is biased warm across much of the domain. This is likely due to the model underestimating mixing, which results in a shoaling of the mixed layer relative to observations. The bias was smallest in the shallow inner domain, which remains well-mixed throughout the water column year round, and largest over deep water, where the model's coarse vertical resolution exacerbates these mixed layer issues.

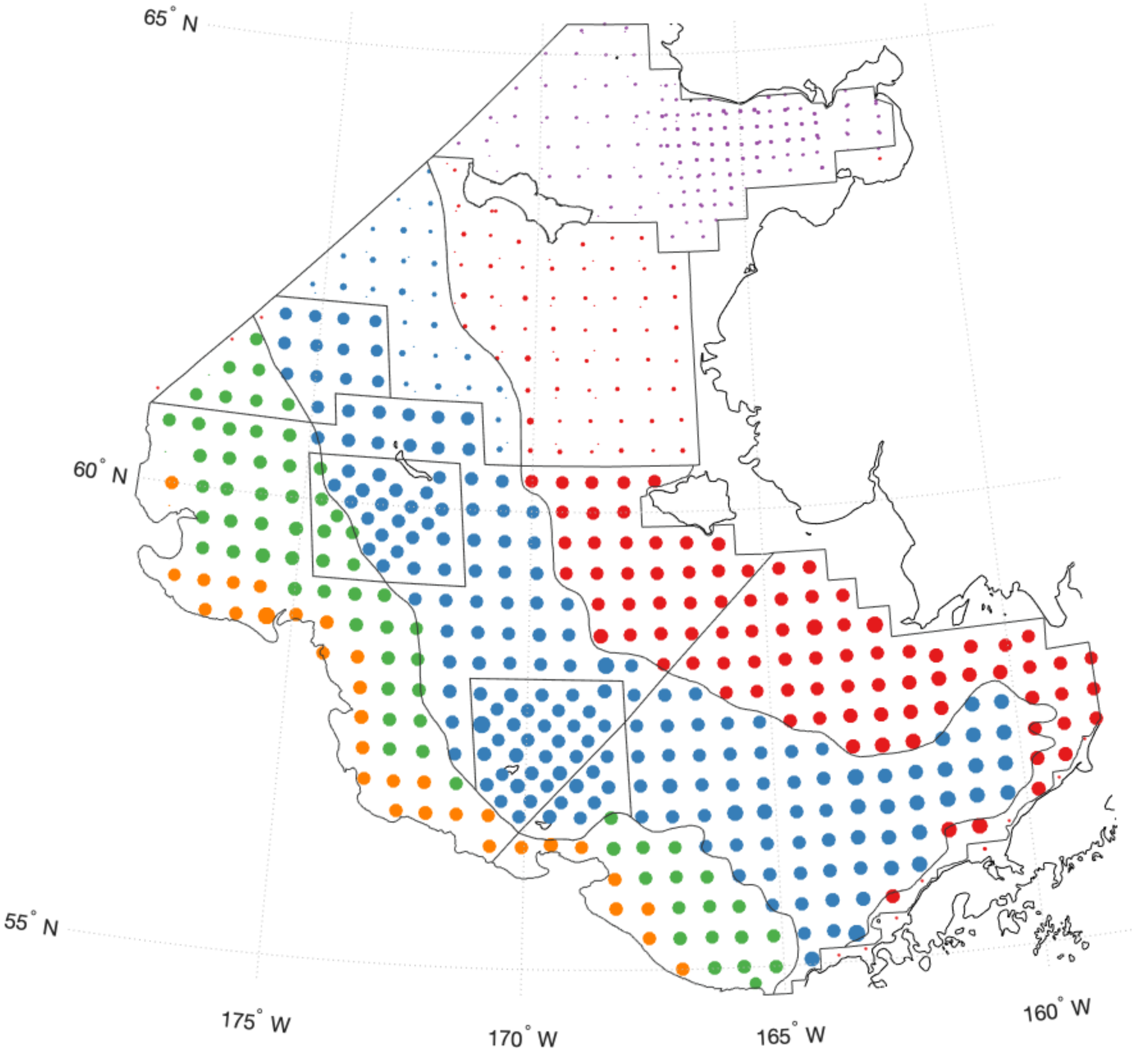

Figure 7. –– A map of analysis regions, with points colored by region and scaled by number of samples. Red indicates the inner shelf, blue the middle shelf, green the outer shelf, orange the shelf break, and purple the northern/Norton Sound region.

Table 3. –– Skill statistics by region, applied to regionally-averaged annual timeseries. Statistics follow Stow et al. (2009), where SD is standard deviation, r is correlation, RMSD is root mean squared difference, cRMSD is centered RMSD, nSD is normalized standard deviation, AAE is average absolute error, and MEF is model efficiency.

Variable	Simulation	Region	SD	r	RMSD	cRMSD	Bias	nSD	AAE	MEF
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Inner	1.376	0.898	0.687	0.607	−0.321	1.100	0.580	0.699
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Middle	0.926	0.931	0.400	0.385	−0.108	0.883	0.334	0.855
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Outer	0.553	0.768	0.383	0.383	0.009	0.969	0.315	0.549
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Northern/Norton	1.990	0.888	1.344	0.955	−0.946	1.327	1.100	0.196
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Shelf break	0.382	0.274	0.734	0.494	0.542	0.875	0.629	−1.829
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	SEBS	0.863	0.910	0.400	0.382	−0.119	0.938	0.342	0.811
Bottom temp.	CFSR/CFSv2 (1982-2018)	Inner	1.318	0.928	0.558	0.493	−0.261	1.114	0.433	0.778
Bottom temp.	CFSR/CFSv2 (1982-2018)	Middle	0.874	0.969	0.314	0.278	−0.147	0.855	0.250	0.905
Bottom temp.	CFSR/CFSv2 (1982-2018)	Outer	0.502	0.674	0.437	0.432	−0.071	0.898	0.360	0.388
Bottom temp.	CFSR/CFSv2 (1982-2018)	Northern/Norton	1.885	0.945	0.785	0.658	0.428	1.217	0.583	0.743
Bottom temp.	CFSR/CFSv2 (1982-2018)	Shelf break	0.461	0.103	0.609	0.599	0.113	1.067	0.499	−0.988
Bottom temp.	CFSR/CFSv2 (1982-2018)	SEBS	0.807	0.962	0.295	0.242	−0.169	0.923	0.242	0.886
Bottom temp.	CORE (1982-2003)	Inner	1.262	0.875	0.715	0.614	−0.366	1.209	0.533	0.531
Bottom temp.	CORE (1982-2003)	Middle	0.728	0.947	0.264	0.264	0.008	0.897	0.219	0.894
Bottom temp.	CORE (1982-2003)	Outer	0.489	0.843	0.306	0.285	0.110	0.937	0.248	0.657
Bottom temp.	CORE (1982-2003)	Northern/Norton	1.151	0.901	1.629	0.499	1.551	1.139	1.551	−1.602
Bottom temp.	CORE (1982-2003)	Shelf break	0.318	0.235	0.784	0.430	0.656	0.854	0.694	−3.417
Bottom temp.	CORE (1982-2003)	SEBS	0.683	0.938	0.242	0.242	−0.008	0.994	0.202	0.876
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Inner	2.068	0.900	1.175	0.936	0.710	1.287	0.865	0.465
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Middle	1.512	0.944	1.656	0.498	1.580	1.078	1.580	−0.394
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Outer	0.977	0.919	1.687	0.419	1.634	0.919	1.634	−1.519
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Northern/Norton	1.302	0.887	1.006	0.760	−0.659	0.803	0.923	0.615
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Shelf break	0.869	0.884	2.113	0.458	2.062	0.886	2.062	−3.645
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	SEBS	1.464	0.942	1.517	0.496	1.433	1.124	1.433	−0.354
Surface temp.	CFSR/CFSv2 (1982-2018)	Inner	1.833	0.960	0.639	0.589	0.248	1.243	0.535	0.812
Surface temp.	CFSR/CFSv2 (1982-2018)	Middle	1.423	0.963	1.381	0.384	1.326	1.054	1.326	−0.046
Surface temp.	CFSR/CFSv2 (1982-2018)	Outer	0.908	0.919	1.467	0.409	1.409	0.878	1.409	−1.009
Surface temp.	CFSR/CFSv2 (1982-2018)	Northern/Norton	1.533	0.906	0.767	0.729	−0.237	0.888	0.650	0.803
Surface temp.	CFSR/CFSv2 (1982-2018)	Shelf break	0.842	0.919	1.739	0.377	1.697	0.882	1.697	−2.324
Surface temp.	CFSR/CFSv2 (1982-2018)	SEBS	1.358	0.969	1.161	0.339	1.110	1.083	1.110	0.143
Surface temp.	CORE (1982-2003)	Inner	1.916	0.887	1.061	0.959	0.456	1.444	0.826	0.360
Surface temp.	CORE (1982-2003)	Middle	1.313	0.939	0.937	0.498	0.794	1.286	0.797	0.156
Surface temp.	CORE (1982-2003)	Outer	0.782	0.900	0.929	0.358	0.857	0.958	0.857	−0.293
Surface temp.	CORE (1982-2003)	Northern/Norton	1.712	0.989	0.632	0.568	−0.276	0.778	0.622	0.917
Surface temp.	CORE (1982-2003)	Shelf break	0.741	0.880	1.474	0.365	1.428	0.989	1.428	−2.866
Surface temp.	CORE (1982-2003)	SEBS	1.298	0.940	0.956	0.491	0.820	1.289	0.832	0.100

Table 4. –– Skill statistics by region, applied to all points. Statistics follow Stow et al. (2009), where SD is standard deviation, r is correlation, RMSD is root mean squared difference, cRMSD is centered RMSD, nSD is normalized standard deviation, AAE is average absolute error, and MEF is model efficiency.

Variable	Simulation	Region	SD	r	RMSD	cRMSD	Bias	nSD	AAE	MEF
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Inner	1.376	0.898	0.687	0.607	−0.321	1.100	0.580	0.699
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Middle	0.926	0.931	0.400	0.385	−0.108	0.883	0.334	0.855
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Outer	0.553	0.768	0.383	0.383	0.009	0.969	0.315	0.549
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Northern/Norton	1.990	0.888	1.344	0.955	−0.946	1.327	1.100	0.196
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Shelf break	0.382	0.274	0.734	0.494	0.542	0.875	0.629	−1.829
Bottom temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	SEBS	0.863	0.910	0.400	0.382	−0.119	0.938	0.342	0.811
Bottom temp.	CFSR/CFSv2 (1982-2018)	Inner	1.318	0.928	0.558	0.493	−0.261	1.114	0.433	0.778
Bottom temp.	CFSR/CFSv2 (1982-2018)	Middle	0.874	0.969	0.314	0.278	−0.147	0.855	0.250	0.905
Bottom temp.	CFSR/CFSv2 (1982-2018)	Outer	0.502	0.674	0.437	0.432	−0.071	0.898	0.360	0.388
Bottom temp.	CFSR/CFSv2 (1982-2018)	Northern/Norton	1.885	0.945	0.785	0.658	0.428	1.217	0.583	0.743
Bottom temp.	CFSR/CFSv2 (1982-2018)	Shelf break	0.461	0.103	0.609	0.599	0.113	1.067	0.499	−0.988
Bottom temp.	CFSR/CFSv2 (1982-2018)	SEBS	0.807	0.962	0.295	0.242	−0.169	0.923	0.242	0.886
Bottom temp.	CORE (1982-2003)	Inner	1.262	0.875	0.715	0.614	−0.366	1.209	0.533	0.531
Bottom temp.	CORE (1982-2003)	Middle	0.728	0.947	0.264	0.264	0.008	0.897	0.219	0.894
Bottom temp.	CORE (1982-2003)	Outer	0.489	0.843	0.306	0.285	0.110	0.937	0.248	0.657
Bottom temp.	CORE (1982-2003)	Northern/Norton	1.151	0.901	1.629	0.499	1.551	1.139	1.551	−1.602
Bottom temp.	CORE (1982-2003)	Shelf break	0.318	0.235	0.784	0.430	0.656	0.854	0.694	−3.417
Bottom temp.	CORE (1982-2003)	SEBS	0.683	0.938	0.242	0.242	−0.008	0.994	0.202	0.876
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Inner	2.068	0.900	1.175	0.936	0.710	1.287	0.865	0.465
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Middle	1.512	0.944	1.656	0.498	1.580	1.078	1.580	−0.394
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Outer	0.977	0.919	1.687	0.419	1.634	0.919	1.634	−1.519
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Northern/Norton	1.302	0.887	1.006	0.760	−0.659	0.803	0.923	0.615
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	Shelf break	0.869	0.884	2.113	0.458	2.062	0.886	2.062	−3.645
Surface temp.	adjusted-CORE/CFSR/CFSv2 (1982-2019)	SEBS	1.464	0.942	1.517	0.496	1.433	1.124	1.433	−0.354
Surface temp.	CFSR/CFSv2 (1982-2018)	Inner	1.833	0.960	0.639	0.589	0.248	1.243	0.535	0.812
Surface temp.	CFSR/CFSv2 (1982-2018)	Middle	1.423	0.963	1.381	0.384	1.326	1.054	1.326	−0.046
Surface temp.	CFSR/CFSv2 (1982-2018)	Outer	0.908	0.919	1.467	0.409	1.409	0.878	1.409	−1.009
Surface temp.	CFSR/CFSv2 (1982-2018)	Northern/Norton	1.533	0.906	0.767	0.729	−0.237	0.888	0.650	0.803
Surface temp.	CFSR/CFSv2 (1982-2018)	Shelf break	0.842	0.919	1.739	0.377	1.697	0.882	1.697	−2.324
Surface temp.	CFSR/CFSv2 (1982-2018)	SEBS	1.358	0.969	1.161	0.339	1.110	1.083	1.110	0.143
Surface temp.	CORE (1982-2003)	Inner	1.916	0.887	1.061	0.959	0.456	1.444	0.826	0.360
Surface temp.	CORE (1982-2003)	Middle	1.313	0.939	0.937	0.498	0.794	1.286	0.797	0.156
Surface temp.	CORE (1982-2003)	Outer	0.782	0.900	0.929	0.358	0.857	0.958	0.857	−0.293
Surface temp.	CORE (1982-2003)	Northern/Norton	1.712	0.989	0.632	0.568	−0.276	0.778	0.622	0.917
Surface temp.	CORE (1982-2003)	Shelf break	0.741	0.880	1.474	0.365	1.428	0.989	1.428	−2.866
Surface temp.	CORE (1982-2003)	SEBS	1.298	0.940	0.956	0.491	0.820	1.289	0.832	0.100

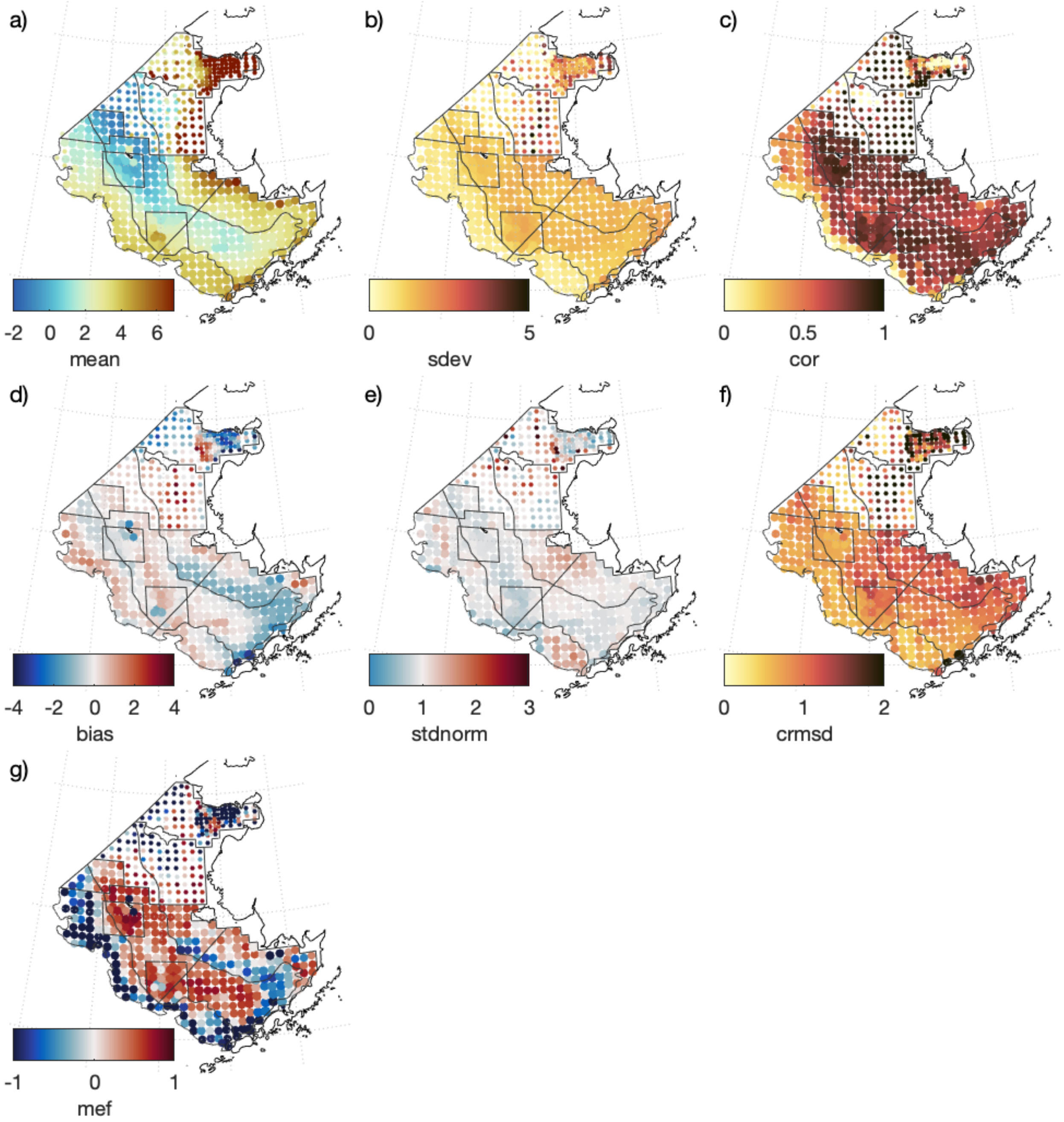

Figure 8. –– Bottom temperature skill by station. Panels a) and b) show the survey data mean and standard deviation, respectively, at each station, with points scaled by the number of times sampled. The remaining panels show bottom temperature skill statistics (after Stow et al. (2009)) for the Bering10K hindcast simulation relative to groundfish survey-derived observations, including c) correlation, d) bias, e) standard deviation in the model relative to that in the observations, f) centered root mean square difference, and g) model efficiency (i.e., skill relative to average of observations, where 1 indicates perfect skill, 0 is as skillful as the mean of observations, and less than 0 is worse than a simple mean predictor).

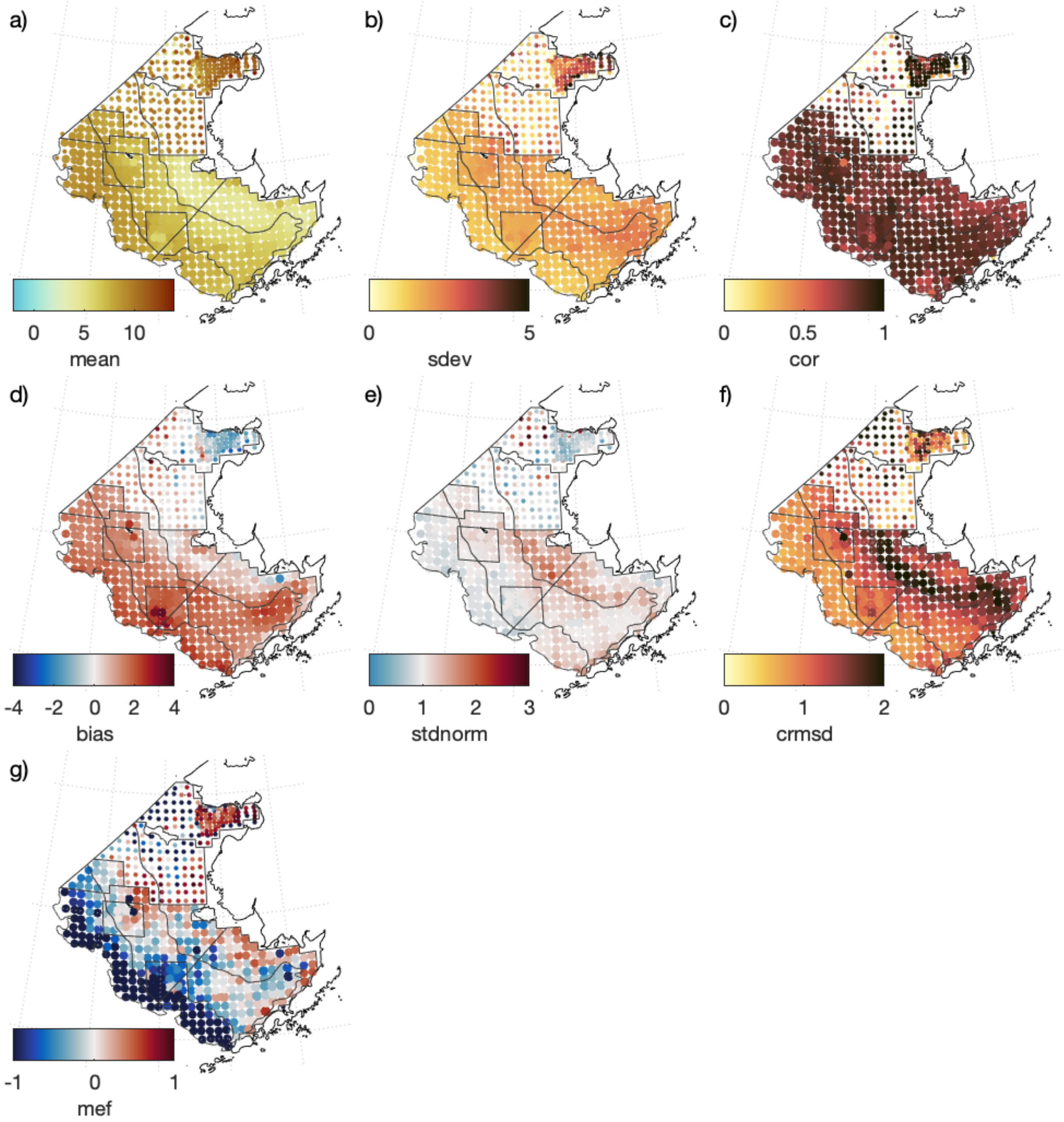

Figure 9. –– Surface temperature skill by station. Panels a) and b) show the survey data mean and standard deviation, respectively, at each station, with points scaled by the number of times sampled. The remaining panels show bottom temperature skill statistics (after Stow et al. (2009)) for the Bering10K hindcast simulation relative to groundfish survey-derived observations, including c) correlation, d) bias, e) standard deviation in the model relative to that in the observations, f) centered root mean square difference, and g) model efficiency (i.e., skill relative to average of observations, where 1 indicates perfect skill, 0 is as skillful as the mean of observations, and less than 0 is worse than a simple mean predictor).

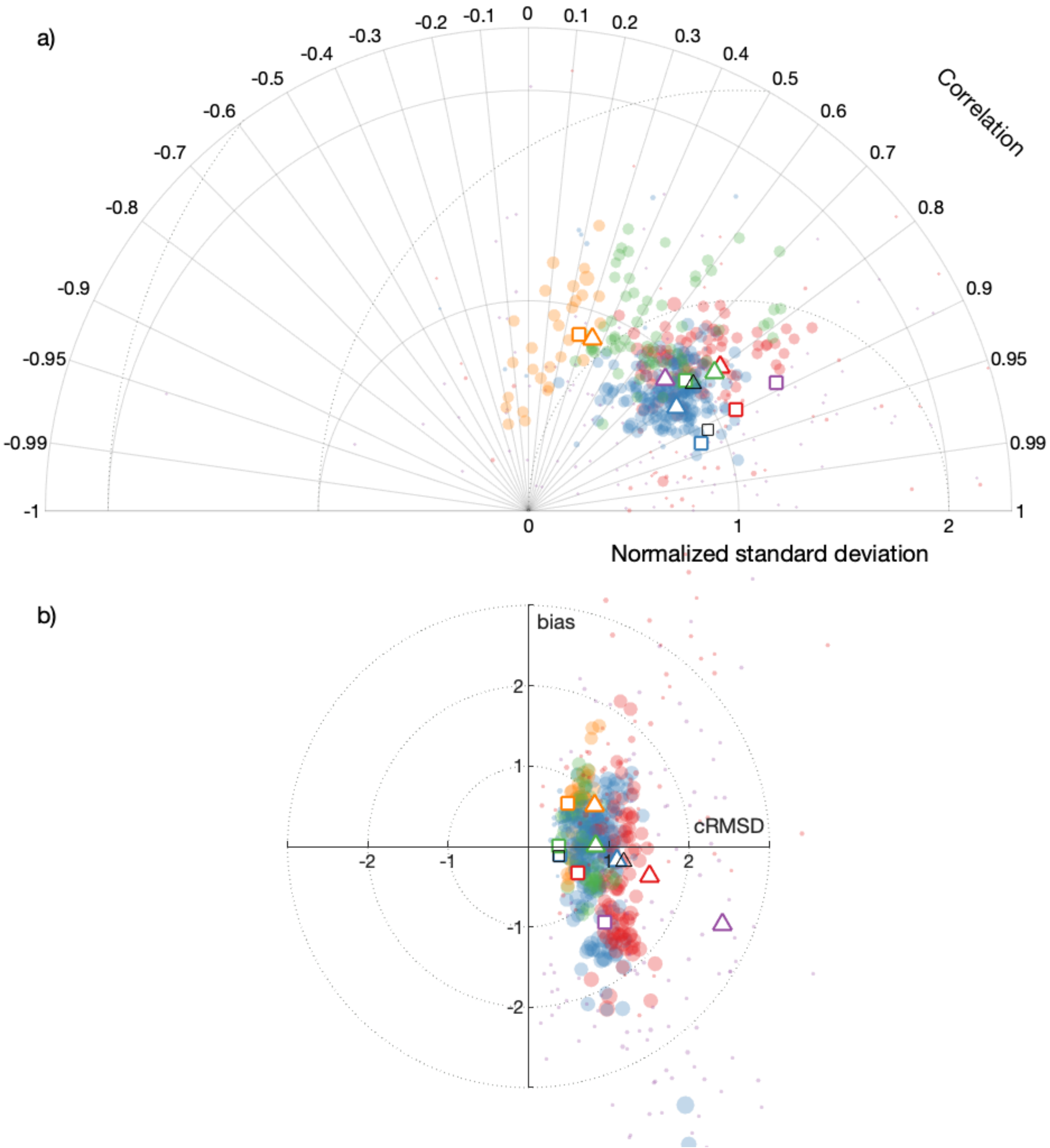

Figure 10. –– Bottom temperature skill statistics, visualized as a) Taylor, and b) target diagrams. Circles indicate each individual station, colored by region (see Fig. 7) and scaled by number of samples, while larger squares and triangles indicate values for the regionally averaged and regionally grouped statistics, respectively. Additional black markers indicate the regional statistics for the SEBS region.

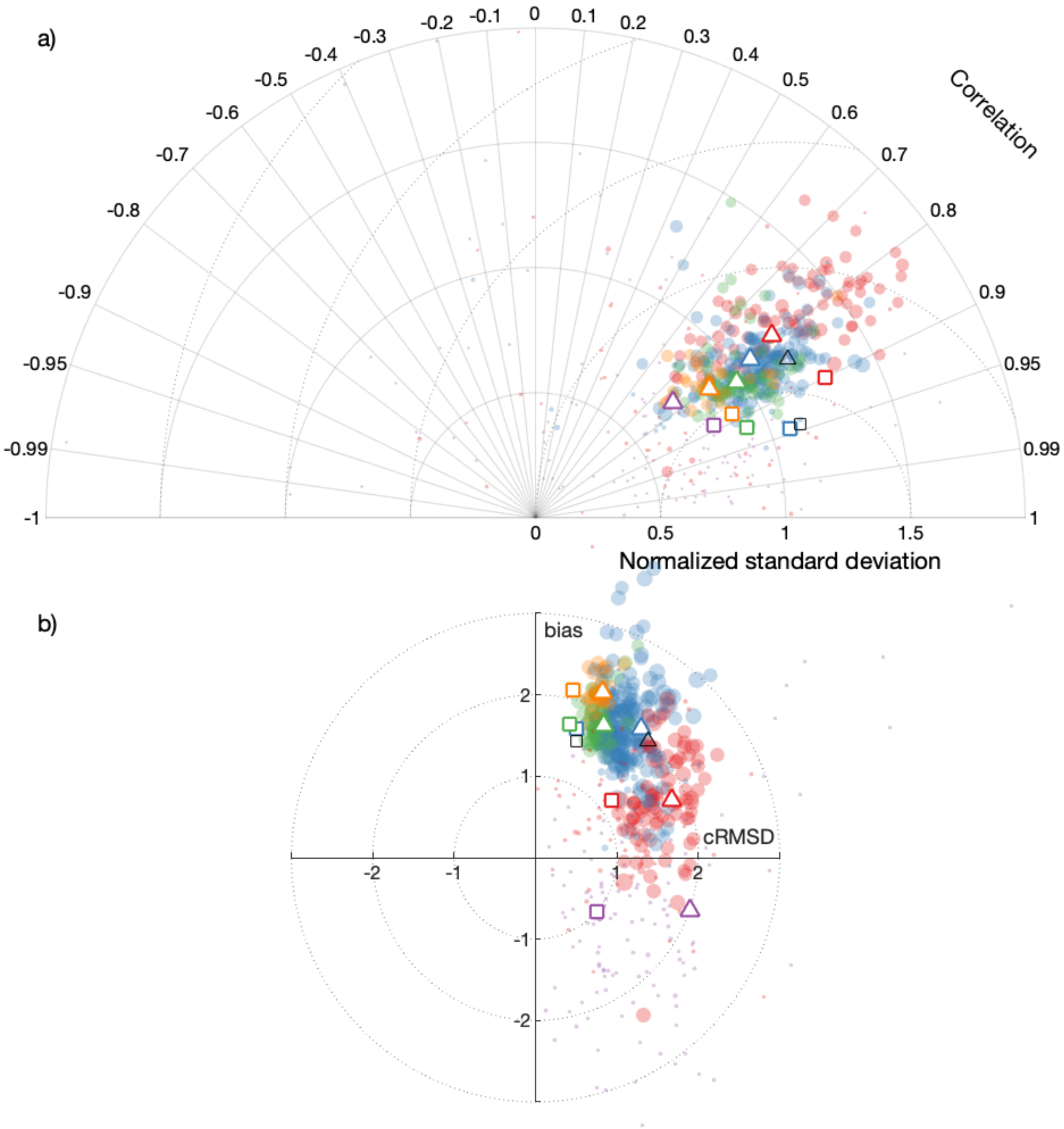

Figure 11. –– Surface temperature skill statistics, visualized as a) Taylor, and b) target diagrams. Circles indicate each individual station, colored by region (see Fig. 7) and scaled by number of samples, while larger squares and triangles indicate values for the regionally averaged and regionally grouped statistics, respectively. Additional black markers indicate the regional statistics for the SEBS region.

Acknowledgments

I would like to thank all those in the Alaska Fisheries Science Center Resource Assessment and Conservation Engineering (RACE) Division who were responsible for collecting, quality-controlling, analyzing, and distributing the survey data from the groundfish surveys. In particular, thanks to Nancy Roberson for running the primary query of the RACEBASE database, and to Lyle Britt for providing some historical context for the dataset. Thanks to Wei Cheng (University of Washington, CICOES) and Kirstin Holsman (NOAA Alaska Fisheries Science Center, Resource Ecology and Fisheries Management Division) for their comments and suggestions to improve the manuscript.

CITATIONS

Buckley, T. W., A. Greig, and J. L. Boldt. 2009. Describing summer pelagic habitat over the continental shelf in the eastern Bering Sea, 1982-2006. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-196, 49 p.

Goddard, P. and M. Zimmerman. 1993. Distribution, abundance, and biological characteristics of groundfish in the eastern Bering Sea based on results of the U.S. bottom trawl survey during June-September 1991. AFSC Processed Report 93-15, 324 p., Alaska Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, 7600 Sand Point Way NE, Seattle, WA 98115-0070.

Haidvogel, D. B., H. Arango, W. P. Budgell, B. D. Cornuelle, E. Curchitser, E. Di Lorenzo, K. Fennel, W. R. Geyer, A. J. Hermann, L. Lanerolle, J. Levin, J. C. McWilliams, A. J. Miller, A. M. Moore, T. M. Powell, A. F. Shchepetkin, C. R. Sherwood, R. P. Signell, J. C. Warner, and J. Wilkin. 2008. Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. J. Comput. Phys. 227(7):3595–3624.

Hermann, A. J., E. N. Curchitser, K. Hedstrom, W. Cheng, N. A. Bond, M. Wang, K. Aydin, P. J. Stabeno, E. D. Cokelet, and G. A. Gibson. 2016. Projected future biophysical states of the Bering Sea. Deep Sea Res. Part II Top. Stud. Oceanogr. 134:30–47.

Hermann, A. J., G. A. Gibson, N. A. Bond, E. N. Curchitser, K. Hedstrom, W. Cheng, M. Wang, P. J. Stabeno, L. Eisner, and K. D. Cieciel. 2013. A multivariate analysis of observed and modeled biophysical variability on the Bering Sea shelf: Multidecadal hindcasts (1970-2009) and forecasts (2010-2040). Deep. Res. Part II Top. Stud. Oceanogr. 94(2011):121–139.

Hollowed, A. B., K. K. Holsman, A. C. Haynie, A. J. Hermann, A. E. Punt, K. Y. Aydin,

J. N. Ianelli, S. Kasperski, W. Cheng, A. Faig, K. Kearney, J. C. P. Reum, P. D. Spencer, I. Spies, W. J. Stockhausen, C. S. Szuwalski, G. Whitehouse, and T. K. Wilderbuer. 2020. Integrated modeling to evaluate climate change impacts on coupled social-ecological systems in Alaska. Front. Mar. Sci. 6(January):1–18.

Holsman, K. K., A. C. Haynie, A. B. Hollowed, J. C. Reum, K. Aydin, A. J. Hermann, W. Cheng, A. Faig, J. N. Ianelli, K. A. Kearney, and A. E. Punt. 2020. Ecosystem-based fisheries management forestalls climate-driven collapse. Nat. Commun. 11(1).

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(2):597—-650.

Large, W. G. and S. G. Yeager. 2009. The global climatology of an interannually varying air–sea flux data set. Clim. Dyn. 33(2):341–364.

Lauth, R. R., E. J. Dawson, and J. Conner. 2019. Results of the 2017 Eastern and Northern Bering Sea Continental Shelf Bottom Trawl Survey of Groundfish and Invertebrate Fauna. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-396, 260 p.

RACE. 2019a. Groundfish survey & species codes. URL https://repository.library.noaa.gov/view/noaa/22662. Accessed October 2020.

RACE. 2019b. Groundfish Survey Vessel & Gear Codes. URL https://repository.library.noaa.gov/view/noaa/22662/noaa_22662_DS2.pdf. Accessed October 2020.

Saha, S., S. Moorthi, H. L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y. T. Hou, H. Y. Chuang, H. M. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. Van Den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J. K. Schemm, W. Ebisuzaki,

R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C. Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg. 2010. The NCEP climate forecast system reanalysis. Bull. Am. Meteorol. Soc. 91(8):1015–1057.

Sample, T. M. and R. J. Wolotira. 1985. Demersal Fish and Shellfish Resources of Norton Sound and Adjacent Waters During 1979. U.S. Dep. Commer., NOAA Tech. Memo. NMFS F/NWC-89, 208 p., NOAA, Natl. Mar. Fish. Serv., NOAA, 7600 Sand Point Way NE, Seattle, WA 98115.

Shchepetkin, A. F. and J. C. McWilliams. 2005. The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Model. 9(4):347–404.

Sigler, M. F., K. Y. Aydin, P. L. Boveng, E. V. F. Jr, R. A. Heintz, and R. R. Lauth. 2015. Alaska Fisheries Science Center Loss of Sea Ice (LOSI) Plan for FY15-FY19. AFSC Processed Rep. 2015-01, 11 p., Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv, 7600 Sand Point Way NE, Seattle, WA 98115.

Stevenson, D. E. and R. R. Lauth. 2012. Latitudinal trends and temporal shifts in the catch composition of bottom trawls conducted on the eastern Bering Sea shelf. Deep. Res. Part II Top. Stud. Oceanogr. 65-70:251–259.

Stevenson, D. E. and R. R. Lauth. 2019. Bottom trawl surveys in the northern Bering Sea indicate recent shifts in the distribution of marine species. Polar Biol. 42(2):407–421.

Stow, C. A., J. Jolliff, D. J. McGillicuddy, S. C. Doney, J. I. Allen, M. A. M. Friedrichs, K. A. Rose, and P. Wallhead. 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76(1-2):4–15.

Wolotira, R. J., T. M. Sample, and M. Morin. 1977. Demersal Fish and Shellfish resources of Norton Sound, the southeastern Chukchi Sea and adjacent waters in the baseline year 1976. NWAFSC Processed Rep. 69 p., Northwest and Alaska Fish. Cent., Natl. Mar. Fish. Serv., NOAA, 7600 Sand Point Way NE, Seattle, WA 98115-0070.