Temperature data from the AFSC Groundfish 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, Joint Institute for the Study of the Atmosphere and Ocean (NOAA Alaska Fisheries Science)

November 2020

Abstract

Blah blah add text

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. Alongside the primary biological measurements (abundance, distribution, 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, this survey-derived temperature data has been used as a primary source of model validation for 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). 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 when preparing the groundfish survey dataset for use in these model-related analyses. It also provides an in-depth analysis of the skill testing that has been performed on the Bering10K model specifically related to this dataset. Finally, it provides a short discussion of the spatiotemporal variability that is present within the survey-derived temperature dataset, and its implications in model comparison and skill assessment.

Groundfish survey data analysis

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 in the south and St. Matthew in the north (Fig. 1, blue circles), encompasses the primary distributional area for commercially important groundfish and shellfish species. This original standard survey region included 329 survey stations arranged on a 20-nautical mile grid. An additional 26 stations (Fig. 1, orange squares) were added at the corner points of the 20-n mi 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 have been systematically sampled via trawl every year. The northern portions of the shelf have been less consistently surveyed. Triennially between 1982 and 1991, trawls were conducted across parts of the northern shelf. FiXme: According to Lyle Britt, these were mostly exploratory for crab... try to find something official about that The northern trawl stations were positioned on a 40-nautical mile grid between St. Matthew and St. Lawrence Islands (Fig. 1, purple +’s), plus a 10-nautical mile 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 expanded to include stratum 81. FiXme: Find out what the 2005-2006 strata 81-ish extra north stuff was about and explain. Possibly following pollock? In 2010, under the AFSC 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 US-Russia Maritime Boundary and using the same 20-nautical mile resolution as in the standard survey area (Fig. 1, brown x’s). 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*). Time constraints prevented a full survey of the northern region to investigate whether fish had migrated north in response to historically low sea ice cover that year; instead, an additional 49 stations on an ad hoc 30-nautical mile grid were added to the 2018 survey (Fig. 1, pink stars).

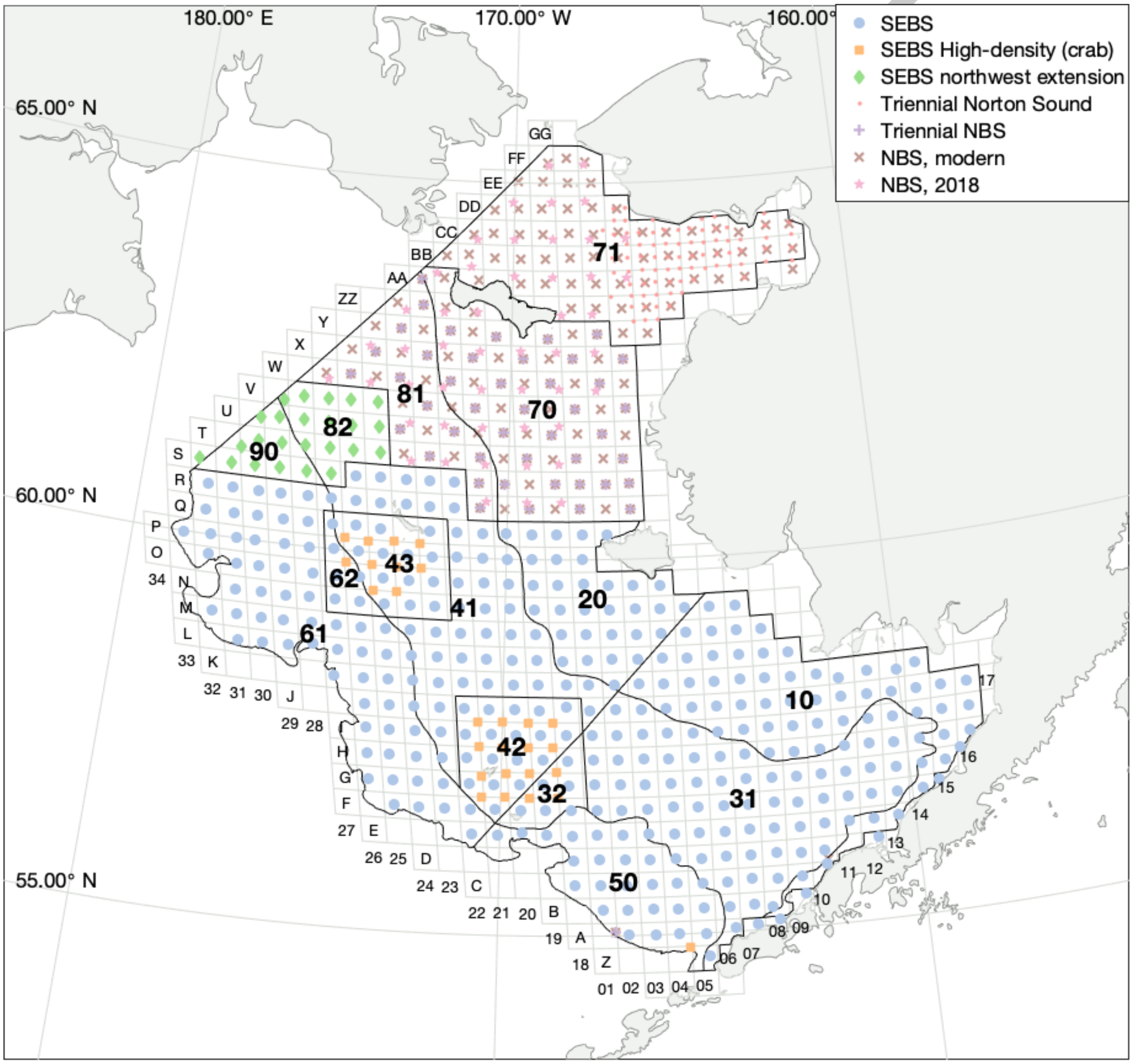

Figure 1. –– Map of Bering Sea groundfish survey sampling sites. The survey strata polygons (black lines, bold numbers) delineate biophysical regions for stratified sampling of the target groundfish species. The primary sampling sites lie on a 20-nautical mile grid (light gray), with each location identified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

The 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. 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.

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 home for the survey data; temperature data was 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, 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 was only available to this author from the latter source.

We began our analysis by combining 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 col-

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
```

lected 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-n mi grid, while in the public dataset they were labeled with the nearest station from the standard 20-n mi grid); in this case we also kept the RACEBASE version only. The public dataset points 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. 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 Bering Sea 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 bathymetric smoothing is necessary to avoid numerical issues in the model, but as a result, direct comparison of modeled bottom temperature to observations becomes more complicated. Therefore, all data collected along the slope was also removed. Finally, in Feb. 1983, a short survey was conducted with samples collected in Bristol Bay, near the southern shelf break, and near the Pribilofs. 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 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 were labeled with station IDs from the 20-n mi. grid, with each point assigned to the station closest to its collection point. However, the samples collected within Norton Sound were actually collected based on a 10-n mi grid. We were unable to locate any description of this grid and the target sample locations associated with it. 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 more or less on a 10-n mi. grid (with wider spacing further 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 indices 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-n mi 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 2-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 US/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; following this pass, missing stratum values were again added based on shared station IDs. Once all entries were assigned a station ID and stratum value, a pass was made to check for inconsistent values. Data from station W-22 was 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 NBS 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 want 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 remains valid despite unsatisfactory performance of the tow for groundfish sampling purposes). See Table **??** for a full description of the variables in the final dataset. This dataset is available in the accompanying AFSC_groundfish_survey_tempe
2020.xlsx spreadsheet 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-n mi. grid, the 10-n mi Norton Sound grid, or the 30-n mi 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 Station-Summary 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 FiXme: DS1 and FiXme: DS2.

<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>R/V Chapman</td></tr>
<tr><td>19</td><td>M/V Pat San Marie</td></tr>
<tr><td>21</td><td>R/V Miller Freeman</td></tr>
<tr><td>37</td><td>R/V Alaska</td></tr>
<tr><td>57</td><td>Morning Star</td></tr>
<tr><td>60</td><td>Argosy</td></tr>
<tr><td>78</td><td>Ocean Hope 3</td></tr>
<tr><td>87</td><td>Tracy Anne</td></tr>
<tr><td>88</td><td>F/V Arcturus</td></tr>
<tr><td>89</td><td>F/V Aldebaran</td></tr>
<tr><td>94</td><td>F/V Vesteraalen</td></tr>
<tr><td>134</td><td>Northwest Explorer</td></tr>
<tr><td>162</td><td>Alaska Knight</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 tracklined 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 near shore 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 satifactory performance, and negative codes indicate unsatisfactory performance. See Appendix i in FiXme: add DS1 citation 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>Longtitude 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 FiXme: others not in DS1?</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></td></tr>
<tr><td>11</td><td></td></tr>
<tr><td>12</td><td></td></tr>
<tr><td>30</td><td></td></tr>
<tr><td>31</td><td></td></tr>
<tr><td>49</td><td></td></tr>
<tr><td>51</td><td></td></tr>
<tr><td>52</td><td></td></tr>
<tr><td>54</td><td></td></tr>
<tr><td>55</td><td></td></tr>
<tr><td>59</td><td></td></tr>
<tr><td>62</td><td></td></tr>
<tr><td>63</td><td></td></tr>
<tr><td>72</td><td></td></tr>
<tr><td>74</td><td></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 FiXme: DS1 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 = 54.64’
	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’ (16.48M) and vertical opening = 3.0M.
	42	Same as 30, but path width = 54.71’ (16.67M) in depths less than 100m.
	43	Same as 30, but path width = 58.41’ (17.80M) in depths greater than 100m.
	44	Same as 30. Acoustic net mensuration equipment attached.
	45	400 Mesh 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-meter beam trawl
ACCESSORIES		

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>2</td><td>6'X 9' steel v-doors, 25fm dandylines branching to 15fm bridle. 1.25" Codend liner, no chains.</td></tr>
<tr><td>15</td><td>6'X 9' steel v-doors (standardized to 1800 lbs after 1988),double 30 fm 5/8" dandylines, 1.28" Mesh codend liner, 24" chain extension between lower dandyline and footrope.</td></tr>
<tr><td>34</td><td>5'X 7' steel vdoors, 25fm dandylines (15fm single, 10fm double), 18" x 8" floats on headrope, 1.25" Mesh liner in codend, no weight on footrope.</td></tr>
<tr><td>47</td><td>6'X 9' steel vdoors, 40fm dandylines (25fm single, 15fm double), 1.25" Liner, no roller gear.</td></tr>
<tr><td>57</td><td>6'X 9' steel vdoors, 2200lbs each. Three 30fm, 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.</td></tr>
<tr><td>64</td><td>Net rigging consists of triple 180' (54.9 M), 5/8" (1.6 Cm) diameter galvanized wire rope dandylines. Dan- dylines are rigged with 18" and 9" chain extensions to the headrope and side panel attachments respectively. Steel v-doors, 6' x 9' (1.83 X 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 x 9 galvinized 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. Ploypropylene 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.</td></tr>
<tr><td>140</td><td>Beam trawl. 3" Pipe frame with semicircle 3" flat strap end runners. 7 Ft.Wide overall x 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)</td></tr>
</table>
</td></tr>
<tr><td>SUBSAMPLE</td><td>Subsampling method</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>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 estimated by volumetric method. Density Lookup Table 2014 used.</td></tr>
<tr><td>13</td><td>Catch subsampled, without load cell weight of catch. Subsample fraction estimated by volumetric method. Density 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 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 (e.g. A-01)</td></tr>
<tr><td>2</td><td>nbs (e.g. NBS-1)</td></tr>
<tr><td>3</td><td>corner (e.g. AB0102)</td></tr>
<tr><td>5</td><td>nesw (e.g AB-S)</td></tr>
<tr><td>6</td><td>norton (e.g. norton01)</td></tr>
<tr><td>8</td><td>sp (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 n-mi grid), corner (of 20 n-mi grid), nbs (30 n-mi grid), or norton (10 n-mi 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

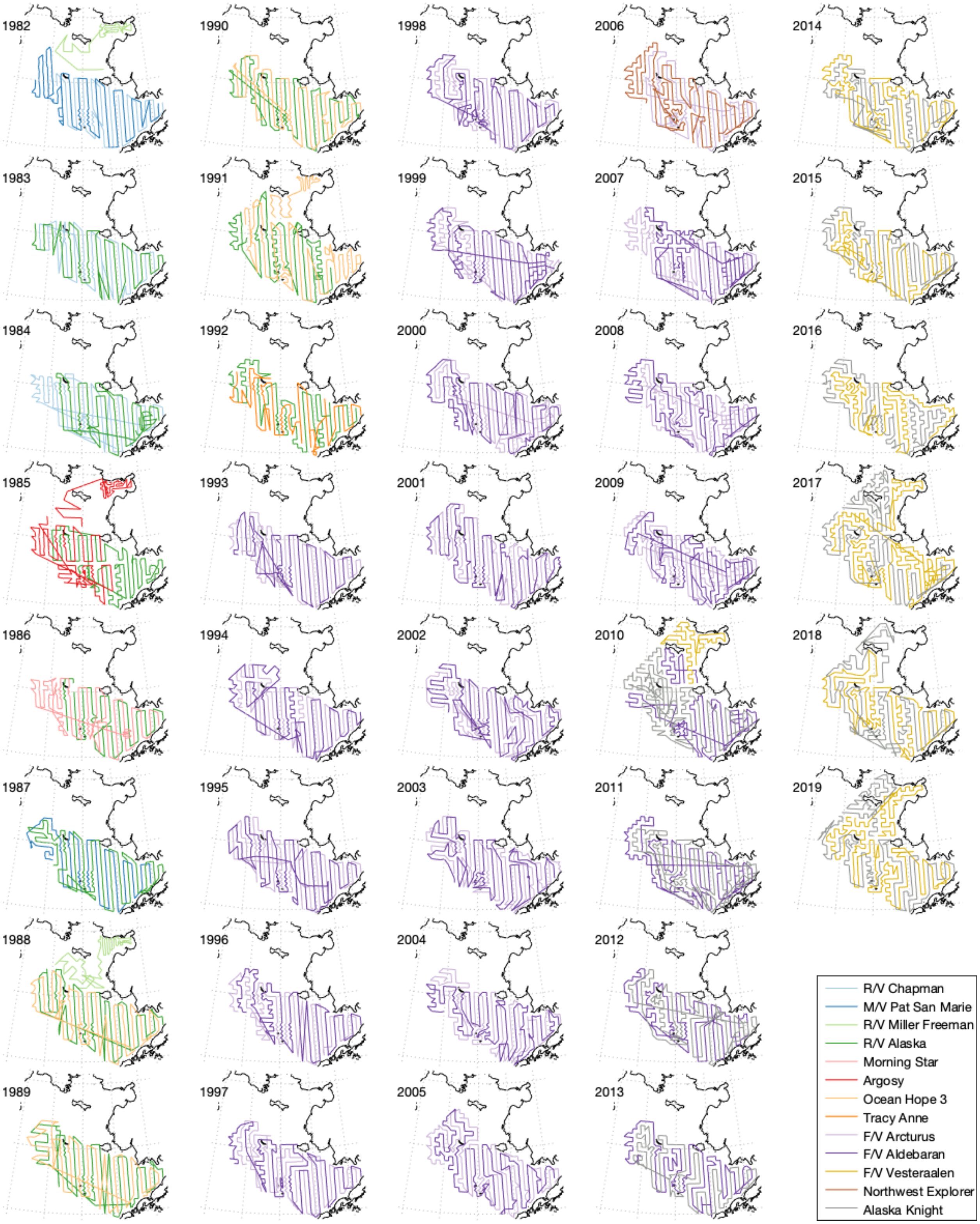

Figure 2. –– Survey cruise tracks by year, colored by survey vessel.

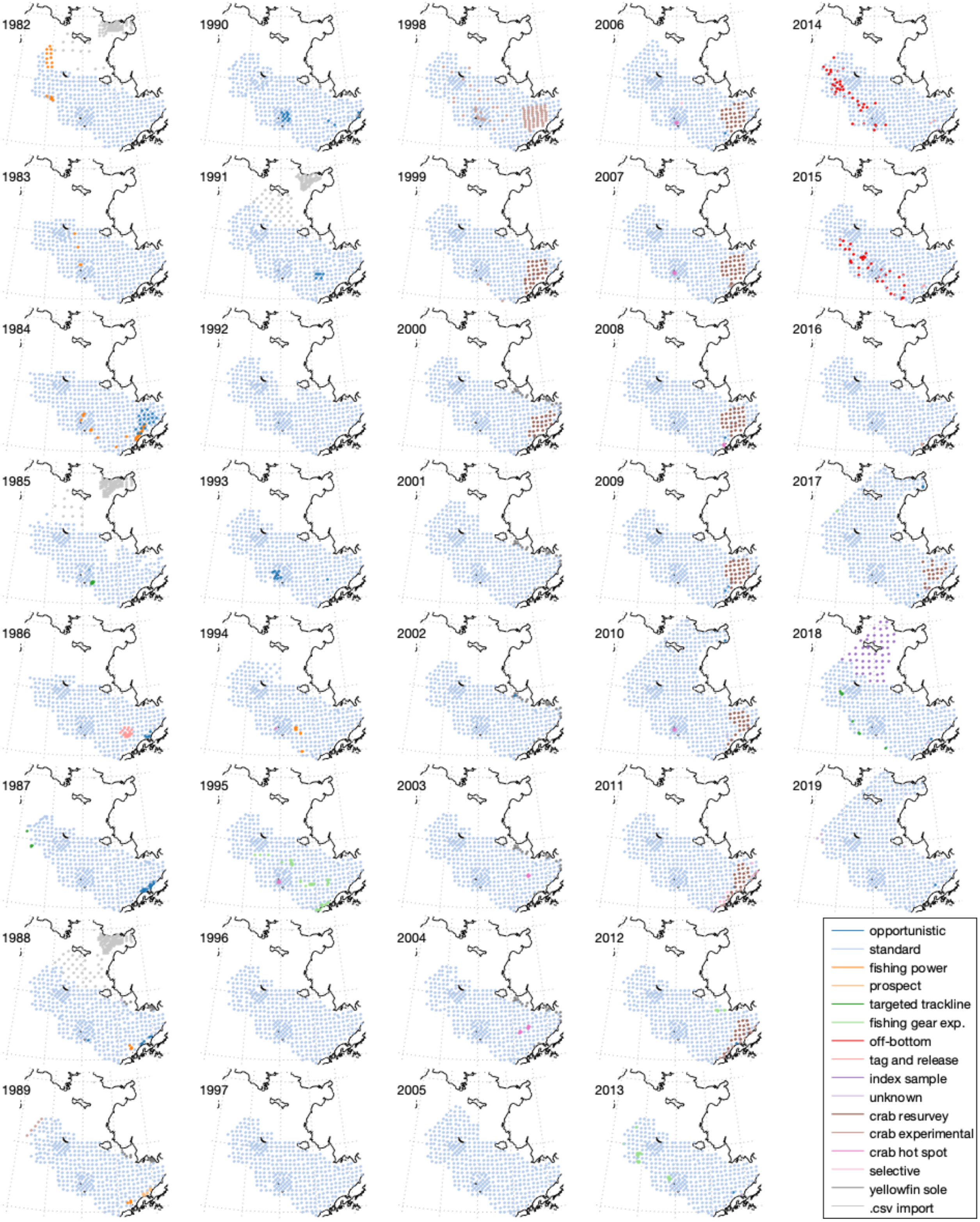

Figure 3. –– Survey sampling sites by year, categorized by haul type.

Resampling a model for comparison with survey dataset

Survey replication methods

We use two common 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 time period as the trawl survey itself, i.e. 1982-2019; 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 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 (for better or worse), 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.

Spatiotemporal variability in the groundfish survey data

Over the summer time period when the survey is conducted, the middle and outer shelf regions (Fig. 1, strata 31–62 and 81–90) are strongly thermally stratified. ~~Because the deeper waters are isolated from surface heating, the~~ bottom temperatures in this region remain relatively constant over the entire survey period (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. 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 sampling-time-derived 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-trawl-derived bottom temperatures from 2010 (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. June 7 and Aug 10, 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).

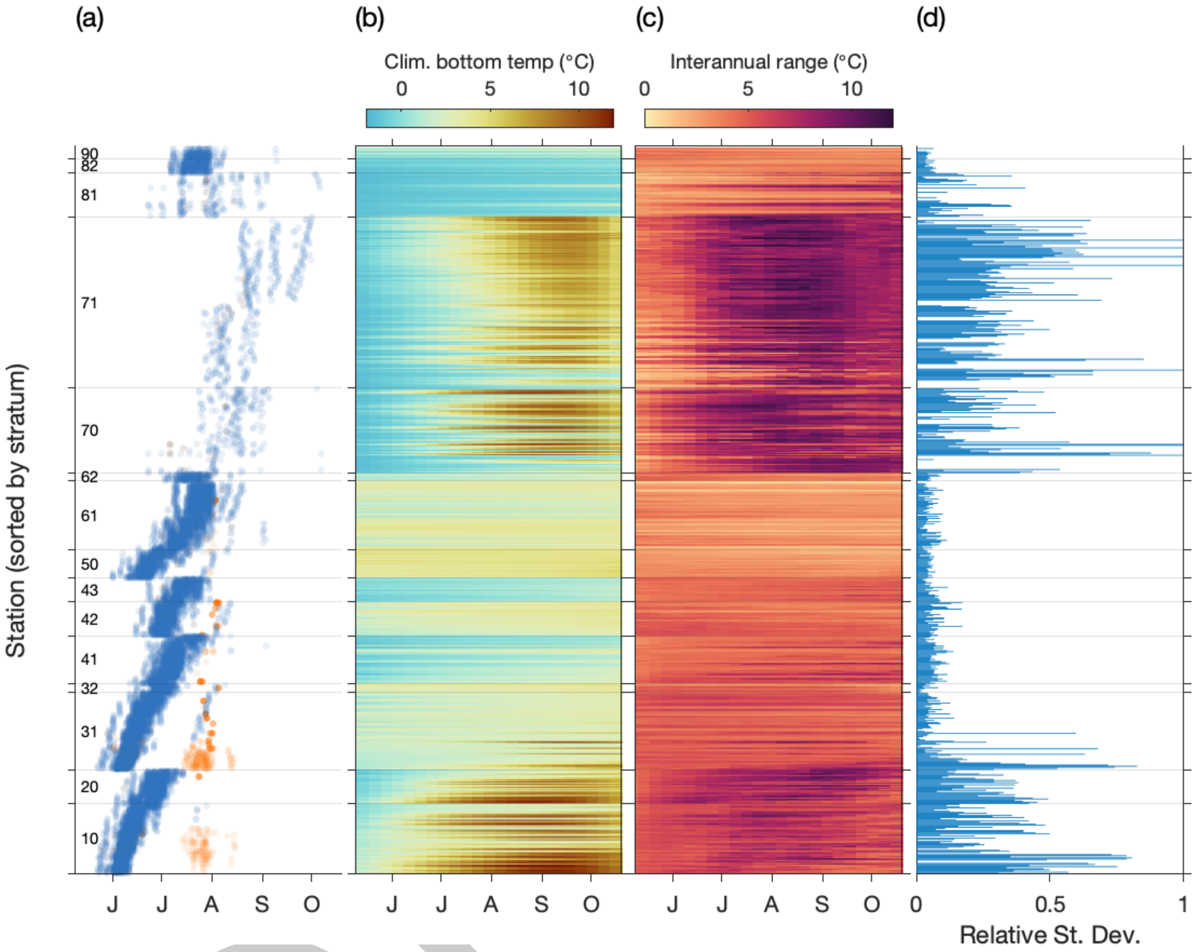

Figure 4. –– Variations in bottom temperature relative to sampling date across the Bering Sea survey region. a) Dots indicate each individual trawl sample, organized by station on the y-axis (sorted by the stratum in which each station is located) and time of year collected on the x-axis. Blue dots indicate the primary trawl at each location, and orange dots indicate additional trawls (incomplete trawls, replications, etc.). b) Climatological hindcasted bottom temperature vs time of year at each station location. c) Interannual range in bottom temperature vs 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 timeseries.

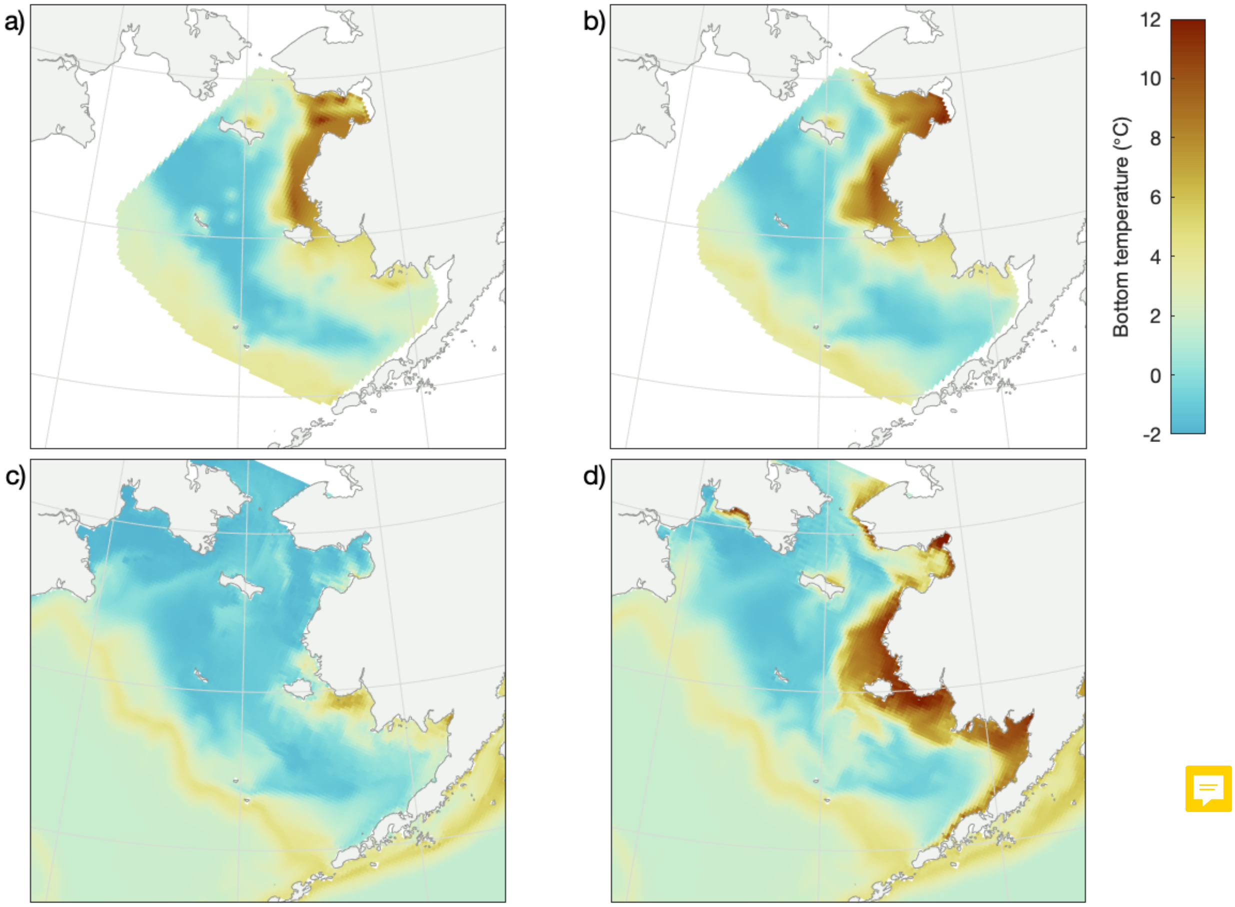

Figure 5. – – a) A composite of bottom temperature based on trawl-sampled survey points from 2010, interpolated to the model grid via a natural neighbor interpolant. b) A composite of bottom temperature based on the Bering10K hindcast sampled using the survey replication method, interpolated to the model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

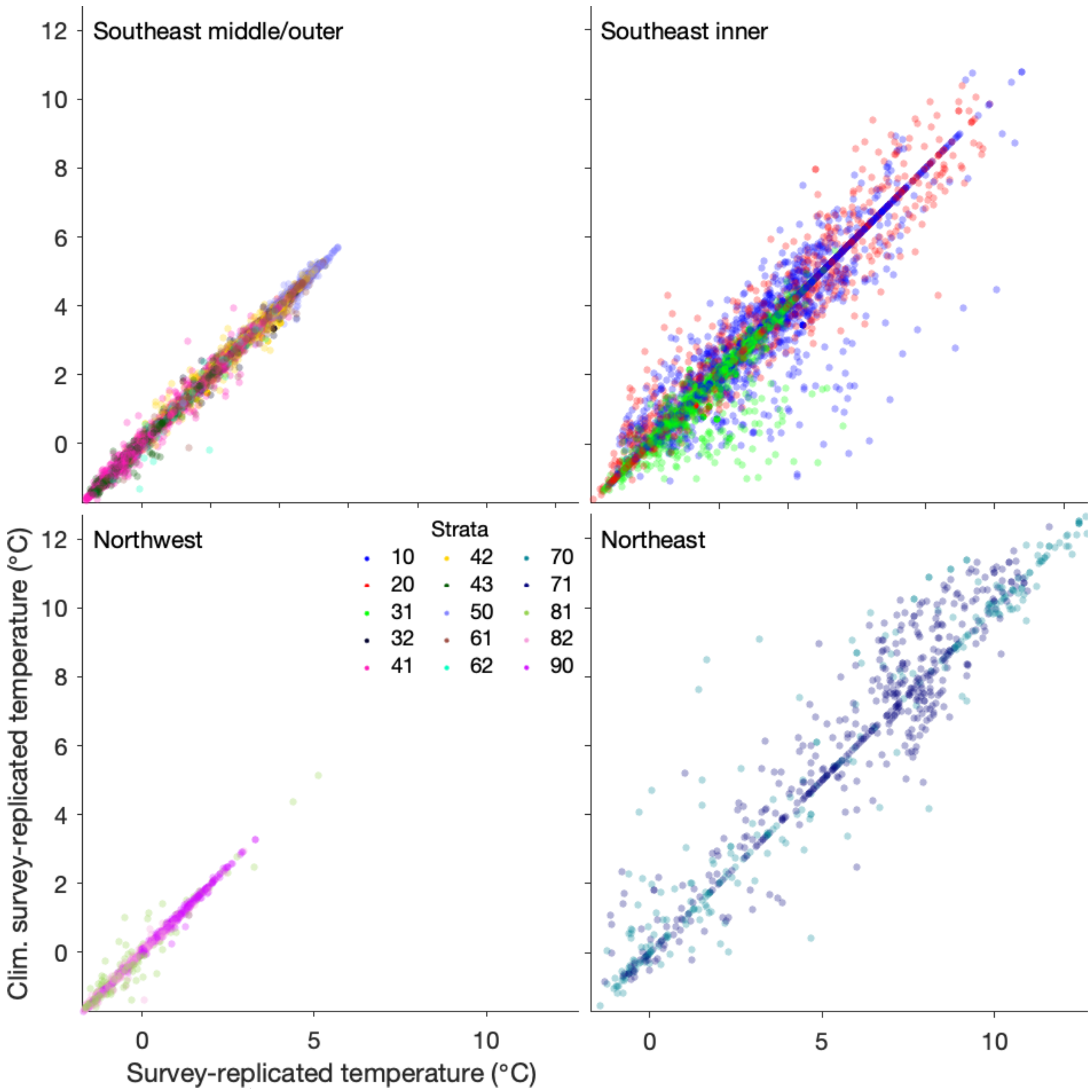

Figure 6. – – A comparsion of survey-replicated bottom temperature extracted from the Bering10K hindcast simulation, compared with an idealized survey-replication of the same simulation, across different regions of the shelf domain.

Bering10K ROMS skill assessment

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 10km horizontal resolution; it was first developed as part of the Bering Ecosystem Study (BEST) and Bering Sea Integrated Ecosystem Research Project (BSIREP), and 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 that 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, but use this report to provide the details underlying them.

The Bering10K hindcast simulations that we evaluate here are 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 includes input forcing from 1970–2003. CFSR spans 1979–March 2011. The CFSv2 Operational Analysis data begins in April 2011, and continues as an operational product to the present. The first hindcast simulation in this study, which is the primary one used for research simulations, uses a combination of these datasets in order to span the longest possible time span. It uses CORE input from 1970–1995, then switches 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 (1970–2003) and just the CFS forcing (1979–2018), without any adjustments to radiation values. The three simulations perform comparably, so we do not include any in-depth analysis of these 3 variants, but do include separate skill statistics 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 on a station by station basis for all stations that had been sampled at least 3 times since 1982 (Fig. 8 and Fig. 9). The stations were then divided into 5 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 6 regions, skill statistics were first calculated on a point-by-point basis, comparing all survey-derived points within each region across time to their survey-replicated counterparts. A separate calculation was performed using regional averages; for this, only the best-replicate points were selected from each year and then averaged into annual time-series for both the survey and survey-replicated datasets (Fig 10, 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 (Table 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.

The station by station skill analysis reveals 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 further inshore than is seen in observations. Skill is 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 performs well across regions in terms of capturing interannual variability. However, 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 is 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.

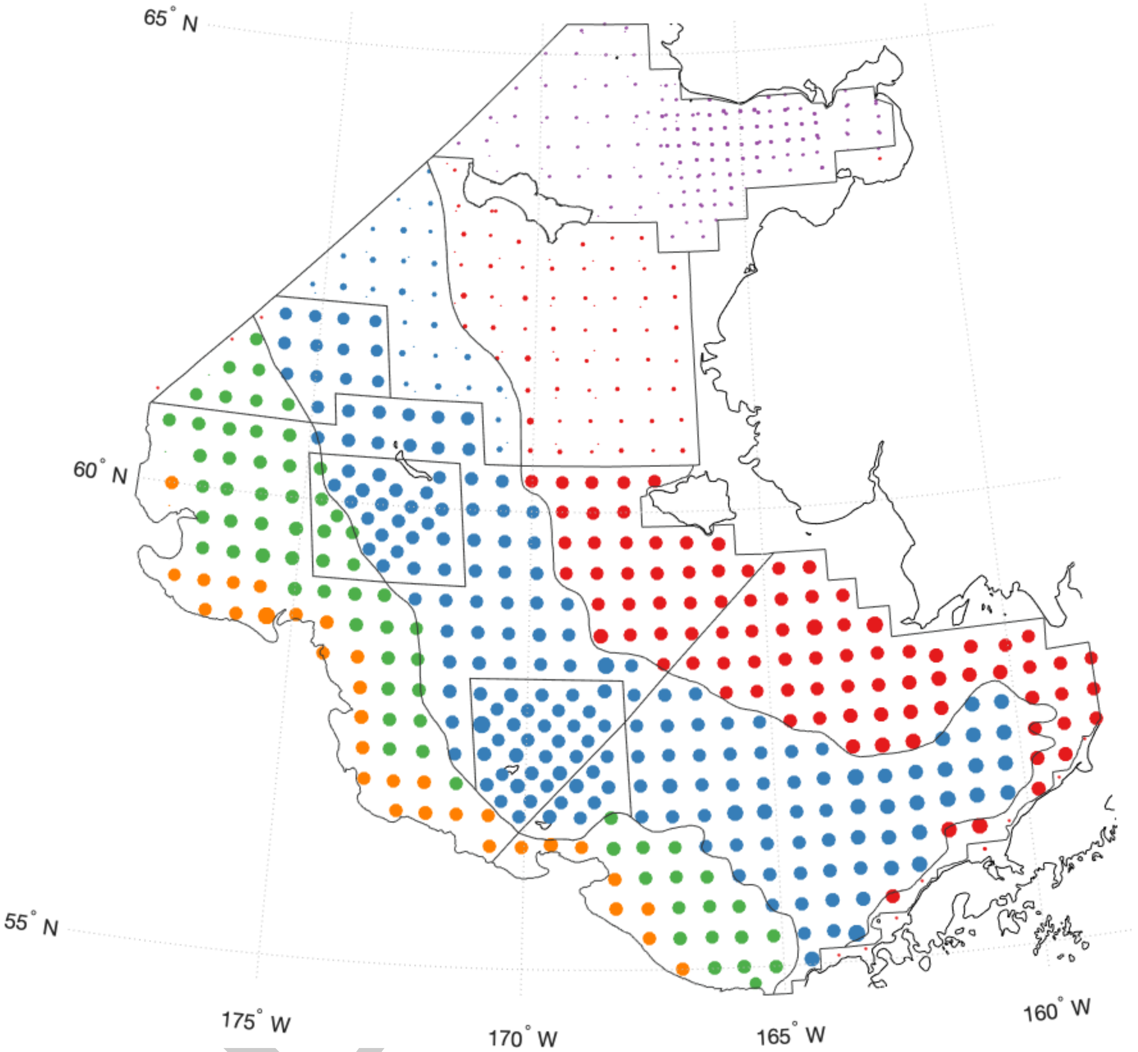

Figure 7. –– A map of analysis regions, with points colored by region and scaled by number of samples. Red indicates the inner shelf, blue the middle shelf, green the outer shelf, orange the shelf break, and purple the northern/Norton Sound region.

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

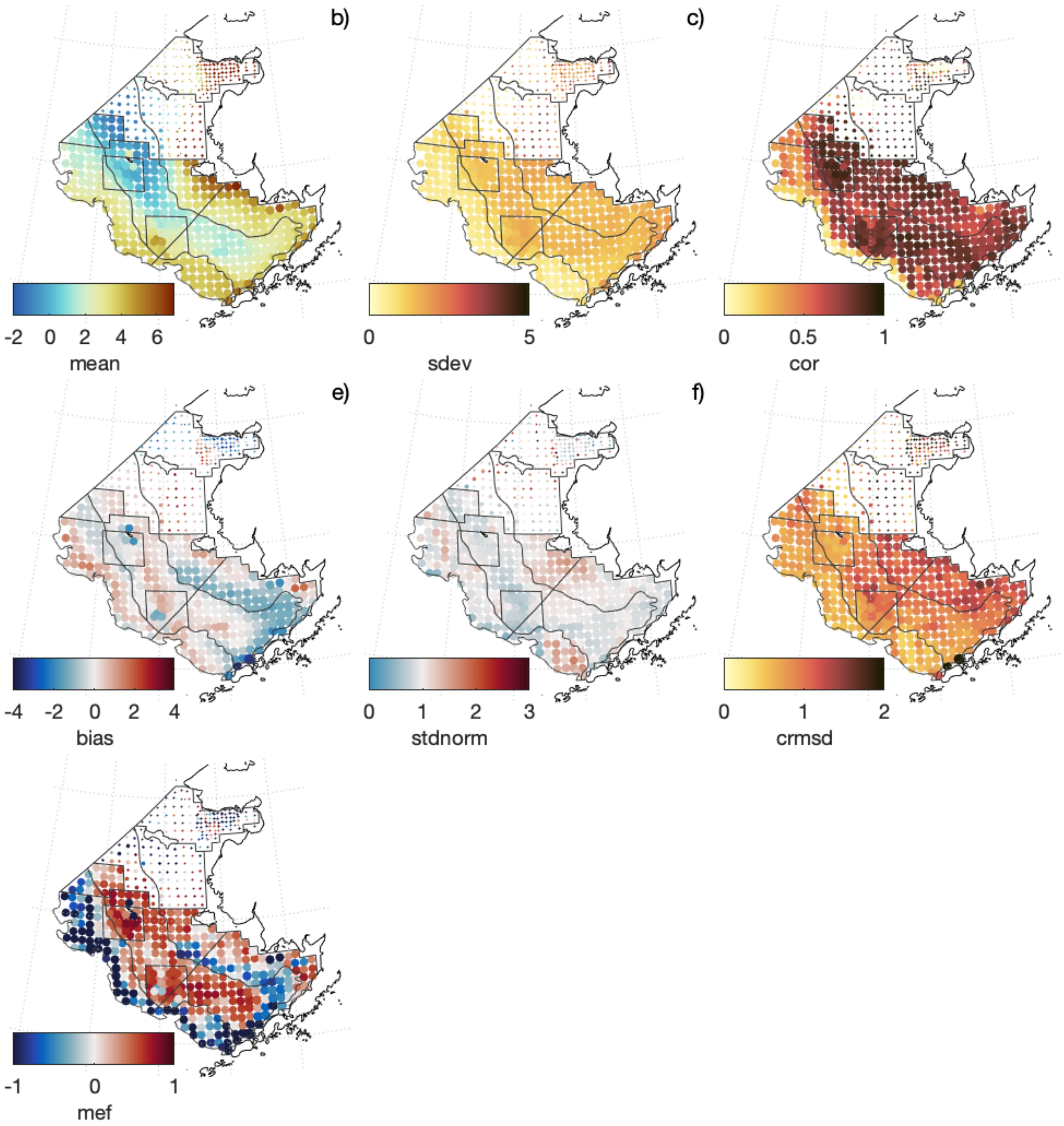

Figure 8. –– Bottom temperature skill by station. Panels a) and b) show the survey data mean and standard deviation, respectively, at each station, with points scaled by the number of times sampled. The remaining panels show bottom temperature skill statistics for the Bering10K hindcast simulation relative to groundfish survey-derived observations, including c) correlation, d) bias, e) normalized standard deviation, i.e. relative to the observations, f) centered root mean square difference, and g) model efficiency.

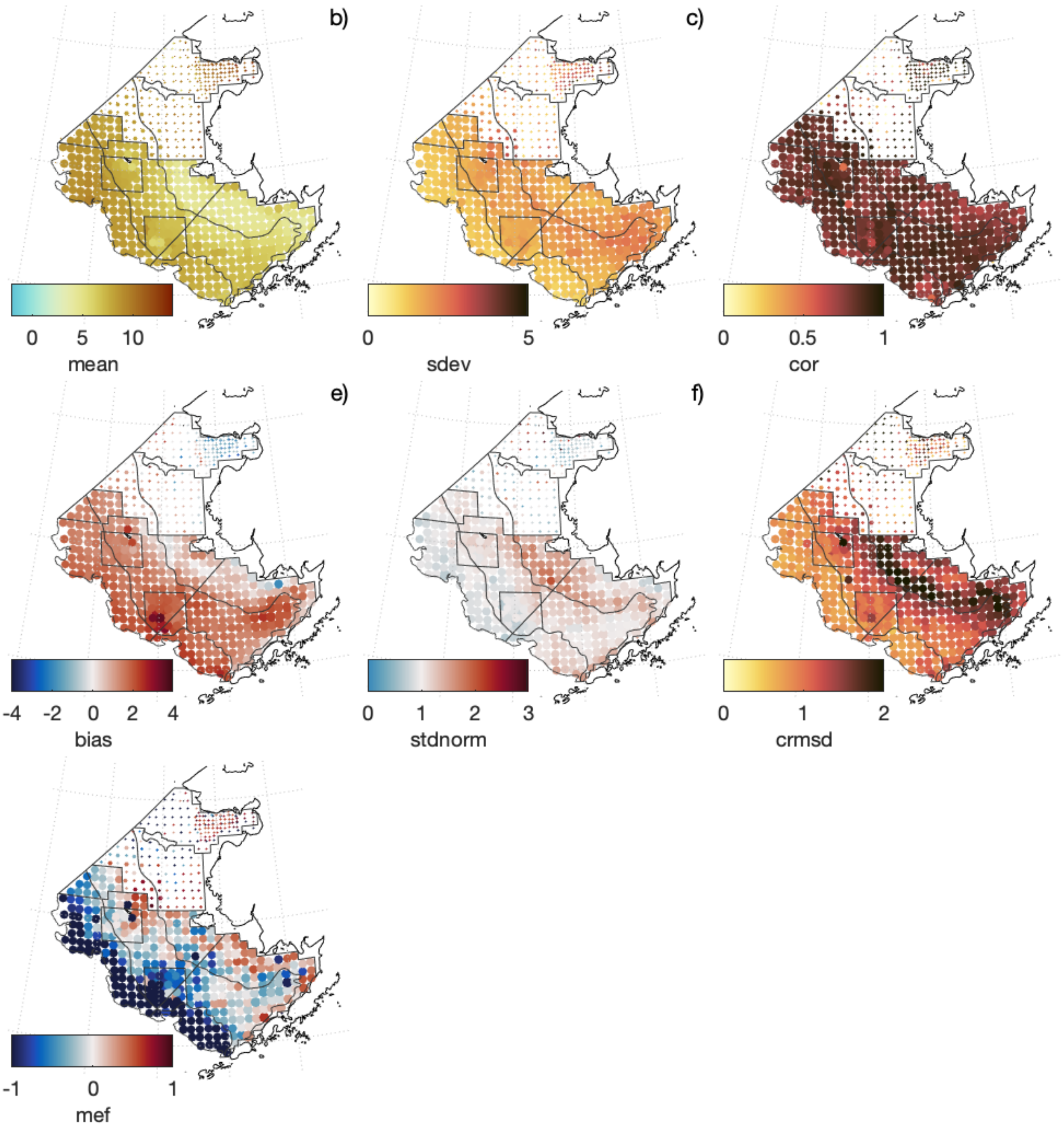

Figure 9. –– Surface temperature skill by station. Panels a) and b) show the survey data mean and standard deviation, respectively, at each station, with points scaled by the number of times sampled. The remaining panels show bottom temperature skill statistics for the Bering10K hindcast simulation relative to groundfish survey-derived observations, including c) correlation, d) bias, e) normalized standard deviation, i.e. relative to the observations, f) centered root mean square difference, and g) model efficiency.

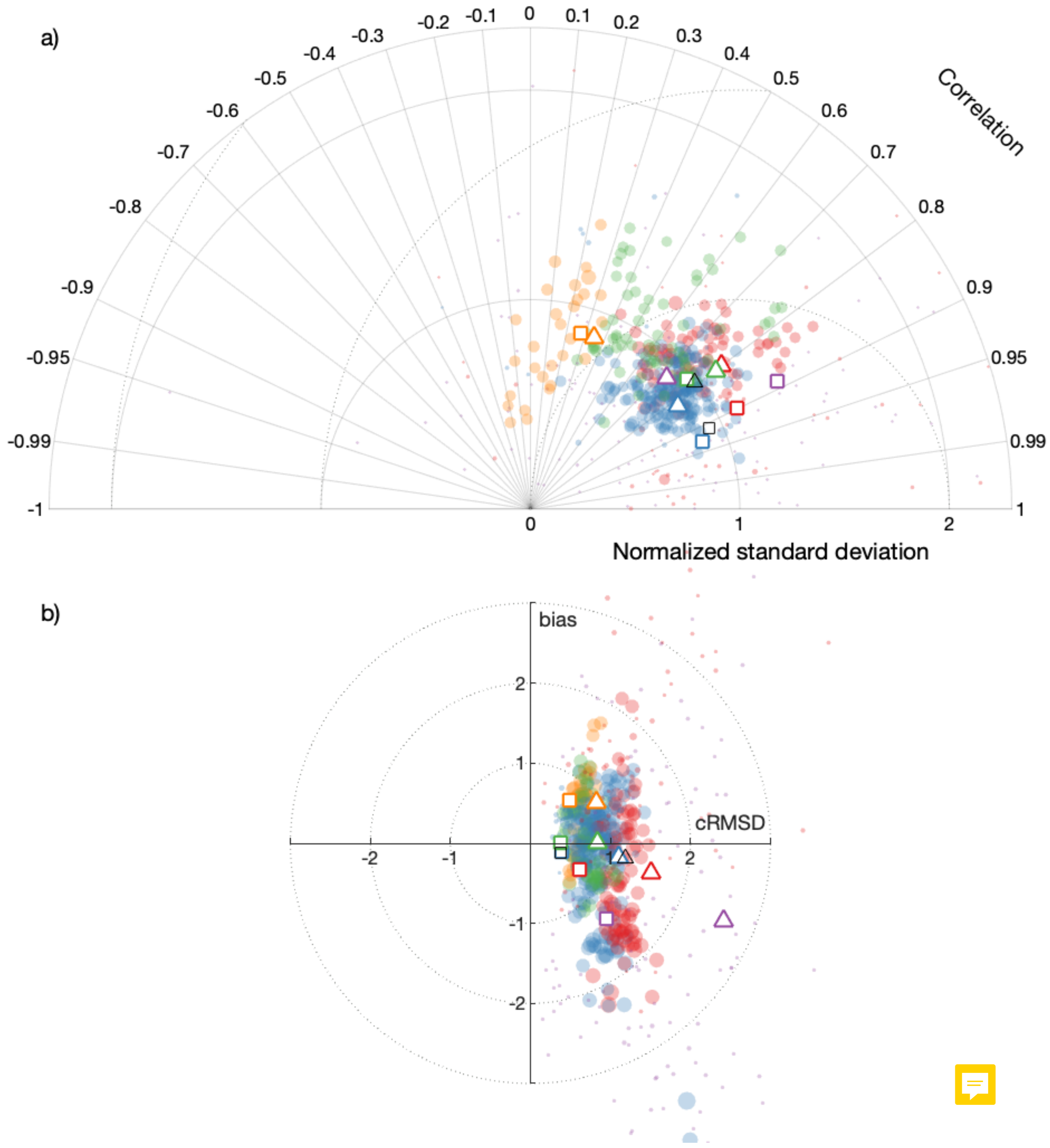

Figure 10. –– Bottom temperature skill statistics, visualized as a) Taylor, and b) target diagrams. Circles indicate each individual station, colored by region (see Fig. 7) and scaled by number of samples, while larger triangles and squares indicate values for the regionally-averaged and regionally-grouped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

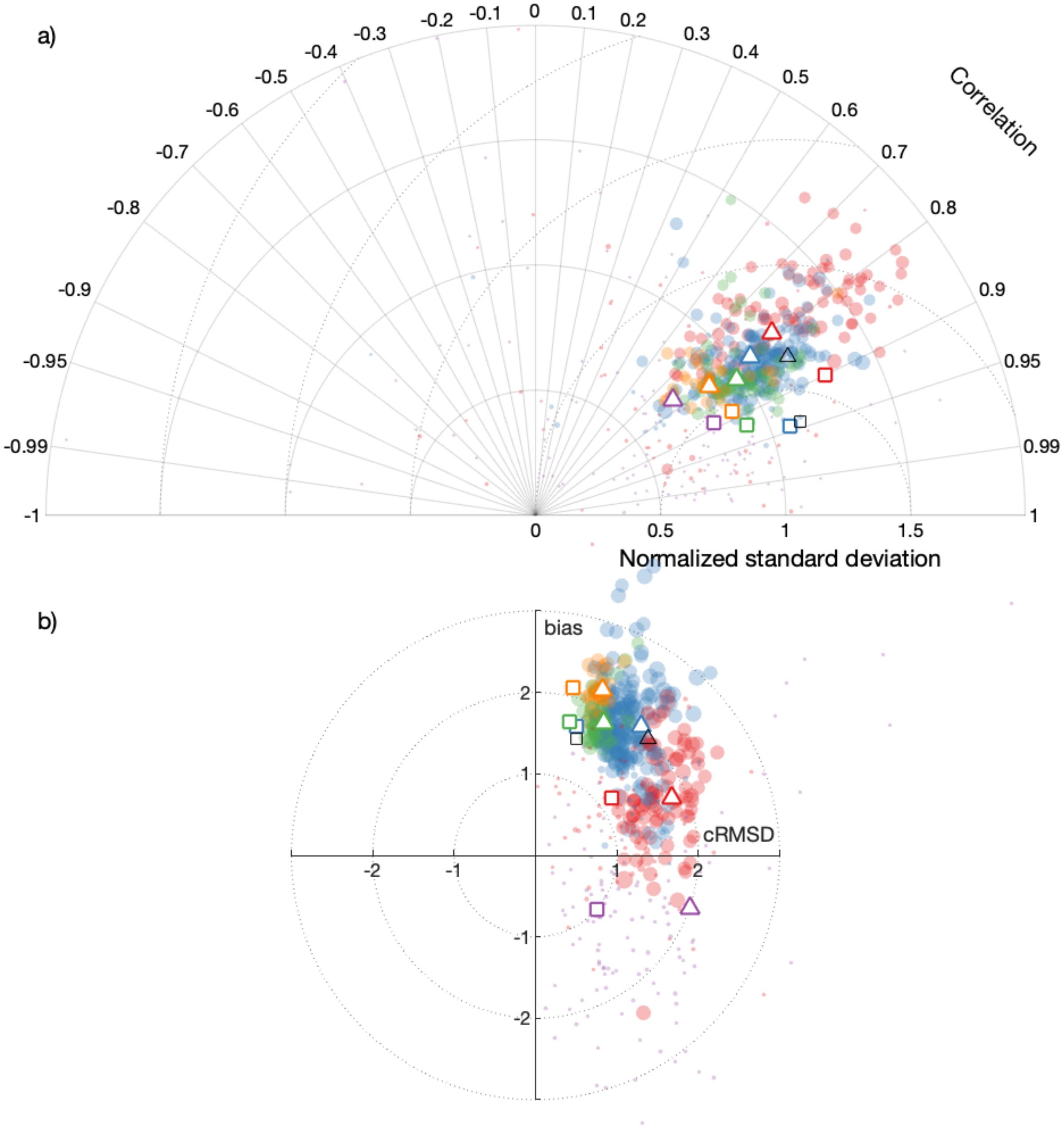

Figure 11. –– Surface temperature skill statistics, visualized as a) Taylor, and b) target diagrams. Circles indicate each individual station, colored by region (see Fig. 7) and scaled by number of samples, while larger triangles and squares indicate values for the regionally-averaged and regionally-grouped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

Citations

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. Journal of Computational Physics 227(7):3595–3624.

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-Sea Research Part II: Topical Studies in Oceanography 94(2011):121–139. URL http://dx.doi.org/10.1016/j.dsr2.2013.04.007.

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. Frontiers in Marine Science 6(January):1–18.

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. Geoscientific Model Development 13(2):597–650. URL https://www.geosci-model-dev.net/13/597/2020/.

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

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. Bulletin of the American Meteorological Society 91(8):1015–1057.

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 Modelling 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 REPORT 2015-01.

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. Journal of Marine Systems 76(1-2):4–15. URL http://dx.doi.org/10.1016/j.jmarsys.2008.03.011.