

# The Bering10K Dataset

Kelly Kearney

March 6, 2021

This document provides an overview of the Bering10K dataset, including an explanation of the simulations and variables available, a description of the model geometry, and usage tips. This is a living document, and it will be updated alongside the dataset as needed.

## Contents

1	<i>Simulations</i>	1
1.1	<i>Projects</i>	1
1.2	<i>File naming scheme</i>	2
1.3	<i>Data access</i>	5
1.4	<i>A few notes and cautions</i>	7
2	<i>The Bering10K coordinate system</i>	8
2.1	<i>The Bering10K grid</i>	8
2.2	<i>Output variable positions</i>	10
2.3	<i>Working with <math>u</math> and <math>v</math></i>	11
2.4	<i>Pre-calculated coordinate values</i>	12
3	<i>The ACLIM indices</i>	12
3.1	<i>Regional average time series</i>	12
3.2	<i>Survey-replicated time series</i>	13
4	<i>Extended grid file regions</i>	14

## 1 Simulations

### 1.1 Projects

This dataset holds output from a variety of simulations of the Bering Sea regional ocean model. The variants include a few different versions of the underlying Bering10K model (i.e. the ocean, sea ice, and biogeochemical module equations and parameters) as well as different input forcing datasets. The simulations can be loosely grouped based on the research project they are part of:

1. **The “operational” hindcast<sup>1</sup>**: A hindcast simulation using the most up-to-date variant of the Bering10K-BESTNPZ code. This simulation

<sup>1</sup> B10K-K20\_CORECFS

begins in 1970 and extends through the present. It is updated approximately three times per year: in Jan, late April, and late August. This simulation is used in a number of different research projects, as well as for ecosystem-based management in the Bering Sea. If you're looking to compare the model to real-world observations, this is probably the simulation you want to use. The data is available on the public server<sup>2</sup>.

2. **ACLIM Phase 1: CMIP5<sup>3</sup>**: These simulations were run as part of the first phase of the Alaska Climate Integrated Modeling (ACLIM) project. They include a hindcast simulation (using an older version of the Bering10K-BESTNPZ model than the operational one) as well as several long-term forecast simulations driven by models from the CMIP5 suite (see Sec. 1.2 for details). The data is available on the public server.
3. **Carbon-enhanced CMIP5 forecasts<sup>4</sup>**: A similar set of simulations as ACLIM Phase 1, but with the a version of the biogeochemical model that includes the carbon system. These data are available to collaborators only.
4. **ACLIM Phase 2: CMIP6<sup>5</sup>**: An update on the ACLIM Phase 1 simulations, using the newer CMIP6 suite for parent models, and with an updated version of the biogeochemical model. These data are available to collaborators only.
5. **Single-parent physics-only hindcasts<sup>6</sup>**: These are physics-only simulations (i.e. biogeochemistry turned off), designed to compare hindcast results using two different parent datasets: CORE2 versus CFSR/CFSv2 (the operational hindcast uses a combination of these two). These hindcasts are used in place of or in addition to the operational hindcast for some research studies. These data are available to collaborators only.

## 1.2 File naming scheme

All output files are named as follows:

```
{ROMS model version}_{parent model/dataset}_{start year}-{end year}_{output type}_{variable}.nc
```

### ROMS model version

All simulations apply a model nickname starting with B10K-, which reflects the model domain (Bering10K, a 10-km grid covering the Bering Sea). The second portion of the name indicates which version of the Bering10K code was used (encompassing source code, input parameters, and vertical resolution):

<sup>2</sup> Live Action Server: <https://data.pmel.noaa.gov/acLim/las/THREDDS>  
Server: <https://data.pmel.noaa.gov/acLim/thredds/>, ERDDAP Server: <https://data.pmel.noaa.gov/acLim/erddap/>

<sup>3</sup> B10K-H16\_CORECFS,  
B10K-H16\_CMIP5\_CESM\_BIO\_rcp85,  
B10K-H16\_CMIP5\_CESM\_rcp45,  
B10K-H16\_CMIP5\_CESM\_rcp85,  
B10K-H16\_CMIP5\_GFDL\_BIO\_rcp85,  
B10K-H16\_CMIP5\_GFDL\_rcp45,  
B10K-H16\_CMIP5\_GFDL\_rcp85,  
B10K-H16\_CMIP5\_MIROC\_rcp45,  
B10K-H16\_CMIP5\_MIROC\_rcp85

<sup>4</sup> B10K-K20P19\_CMIP5\_CESM\_rcp45,  
B10K-K20P19\_CMIP5\_CESM\_rcp85,  
B10K-K20P19\_CMIP5\_GFDL\_rcp45,  
B10K-K20P19\_CMIP5\_GFDL\_rcp85,  
B10K-K20P19\_CMIP5\_MIROC\_rcp45,  
B10K-K20P19\_CMIP5\_MIROC\_rcp85

<sup>5</sup> B10K-K20P19\_CMIP6\_cesm\_historical,  
B10K-K20P19\_CMIP6\_cesm\_ssp126,  
B10K-K20P19\_CMIP6\_cesm\_ssp585,  
B10K-K20P19\_CMIP6\_gfdl\_historical,  
B10K-K20P19\_CMIP6\_gfdl\_ssp126,  
B10K-K20P19\_CMIP6\_gfdl\_ssp585,  
B10K-K20P19\_CMIP6\_miroc\_historical,  
B10K-K20P19\_CMIP6\_miroc\_ssp126,  
B10K-K20P19\_CMIP6\_miroc\_ssp585

<sup>6</sup> B10K-K20nobio\_CFS,  
B10K-K20nobio\_CORE

<sup>7</sup> A. J. Hermann, E. N. Curchitser, K. Hedstrom, W. Cheng, N. A. Bond, M. Wang, K. Aydin, P. J. Stabeno, E. D. Cokelet, and G. A. Gibson. Projected future biophysical states of the Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 134:30–47, 2016. ISSN 09670645. doi: 10.1016/j.dsr2.2015.11.001. URL <http://dx.doi.org/10.1016/j.dsr2.2015.11.001>

- **H16:** 10-layer variant with BESTNPZ biological module circa Hermann et al. [2016]<sup>7</sup>.  
TECHNICAL DETAILS: CMIP5 simulations were run prior to the start of version control. A new executable was later compiled from the ROMS Bering Sea code<sup>8</sup> on 01/30/2018, using the acim branch (commit 2dd1792dfb). In addition to the compilation flags set in the header file (Apps/NEP/nep5.h), the following flags were defined: BEST\_NPZ. This executable was used for the H16 hindcast.
- **K20:** 30-layer variant with the BESTNPZ biological module circa Kearney et al. [2020]<sup>9</sup>.  
TECHNICAL DETAILS: The executable was compiled from the ROMS Bering Sea code<sup>10</sup> on 04/12/2019, using the master branch (commit 0a2d6c80e7). In addition to the compilation flags set in the header file (Apps/NEP/nep5.h), the following flags were defined: BEST\_NPZ, PI\_CONSTANT, and GPPMID.
- **P19:** 10-layer variant, same as H16 but with the addition of carbonate chemistry variables, as documented in [Pilcher et al., 2019]<sup>11</sup>
- **K20P19:** 30-layer variant that merges the biological module source code and parameter updates from the K20 version with the carbonate chemistry updates from the P19 version.  
TECHNICAL DETAILS: The executable for the CMIP6 forecasts was compiled from the ROMS Bering Sea code<sup>12</sup> on 12/19/2019, using the mergetest branch (commit 4d986a7839). In addition to the compilation flags set in the header file (Apps/NEP/nep5.h), the following flags were defined: BEST\_NPZ, PI\_CONSTANT, GPPMID, and CARBON.
- **K20nobio:** uses same source code as K20, but without the BESTNPZ flag turned on, i.e. this is a physics-only simulation.

### Parent model/dataset

Each ROMS simulation represents the downscaling of a coarser-resolution model simulation or reanalysis data product; the parent model/dataset provides surface atmospheric boundary conditions and lateral ocean boundary conditions.

- **CORECFS:** Hindcast simulations. The simulation begins in 1970 and extends through the present (or through what was the present at the time a particular simulation was run). This simulation uses three different datasets for atmospheric surface forcing:
  - 1970-1994: Coordinated Ocean-ice Reference Experiments, version 2 (CORE<sup>13</sup>)
  - 1995-March 2011: Climate Forecast System Reanalysis (CFSR<sup>14</sup>)

<sup>8</sup> <https://github.com/beringnpz/roms-bering-sea>

<sup>9</sup> K. Kearney, A. Hermann, W. Cheng, I. Ortiz, and K. Aydin. 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, 2020. doi: 10.5194/gmd-13-597-2020. URL <https://www.geosci-model-dev.net/13/597/2020/>

<sup>10</sup> <https://github.com/beringnpz/roms-bering-sea>

<sup>11</sup> D. J. Pilcher, D. M. Naiman, J. N. Cross, A. J. Hermann, S. A. Siedlecki, G. A. Gibson, and J. T. Mathis. Modeled Effect of Coastal Biogeochemical Processes, Climate Variability, and Ocean Acidification on Aragonite Saturation State in the Bering Sea. *Frontiers in Marine Science*, 5(January):1–18, 2019. doi: 10.3389/fmars.2018.00508

<sup>12</sup> <https://github.com/beringnpz/roms-bering-sea>

<sup>13</sup> W. G. Large and S. G. Yeager. The global climatology of an interannually varying air–sea flux data set. *Climate dynamics*, 33(2):341–364, 2009. ISSN 0930-7575

<sup>14</sup> S. Saha, 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. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1057, 2010. ISSN 00030007. doi: 10.1175/2010BAMS3001.1

- April 2011-present: Climate Forecast System Operational Analysis (CFSv2)

All datasets are reanalysis products and reflect observed atmospheric conditions.

See Kearney et al. [2020] for further details of model input, including a description of lateral boundary conditions, river input data, and initialization details.

- **CMIP5\_<XXXX>\_rcp<YYY>**: These simulations downscale a selection of the representative concentration pathway simulations from the fifth phase of the Climate Model Intercomparison Project (CMIP5). Names reflect the parent model simulation (MIROC = MIROC-ESM, CESM = CESM1, GFDL = GFDL ESM2M) and concentration pathway (rcp45 = RCP 4.5, rcp85 = RCP 8.5). Some simulations also include a BIO tag in the name; this indicates that these simulations used parent-model-derived lateral boundary conditions for the nutrient variables (NO<sub>3</sub> and NH<sub>4</sub>) as opposed to the climatological values used in all other forecast simulations.
- **CMIP6\_<XXXX>\_ssp<YYY>**: These simulations downscale a selection of the historical and shared socioeconomic pathway simulations from the sixth phase of the Climate Model Intercomparison Project (CMIP6). Names reflect the parent model simulation (miroc = MIROC ES2L, cesm = CESM2, gfdl = GFDL ESM4) and emissions scenario (ssp126 = SSP1-2.6, ssp585 = SSP5-8.5, historical = Historical).

### Start and end year

Simulation output is divided into 5-year time intervals. Files belonging to time blocks at the beginning or end of a simulation may contain less than 5 years worth of data.

### Output type

ROMS model output includes three file types:

- **average**: Average output is archived weekly for the entire grid and reflects weekly-averaged data. Time values corresponding to each datapoint indicate the mid-point of the time-averaged interval.
- **history**: History output is archived weekly for the entire grid and reflects the exact snapshot at the indicated time. Note that certain variables, such as current speeds, may include aliasing due to mismatches between tidal cycles and the weekly archiving interval.
- **stations**: Station output is archived every 6 hours at a few select locations. The station locations were chosen to either correspond

to an observation platform (e.g. M2 and other PMEL moorings) or to be representative of larger biophysical regions. Like history files, these datapoints reflect specific snapshots at the time of archiving and may be subject to aliasing.

## Variable

The variable label in each file name indicates the primary output variable saved in that file. The one exception to this is the constants file, which holds all non-time-dependent variables associated with a simulation.

See the `Bering10K_simulation_variables.xls`<sup>15</sup> spreadsheet for a list of the variables available in each simulation. In the dataset, variables are organized into folders based on the following post-processing levels:

- Level 0: Raw model output files (not included in public dataset)
- Level 1: Model output variables, calculated directly by ROMS. These are stored in netCDF format, and are located on the native model grids. (This is the same data as found in the Level 0 files, but in a more user-friendly format).
- Level 2: Derived output variables, calculated after the simulation is complete based on other Level 1 and Level 2 output variables. These are stored in netCDF format, and are located on the native model grids. These derived output variables are provided to reduce repeated calculation of some commonly-requested quantities (e.g. surface and bottom temperature, depth-integrated versions of biological tracer variables, etc.).
- Level 3: Derived variables, indices, etc., decoupled from the native model grid. File format varies based on contents. Currently, this includes the ACLIM indices time series (see section 3).

## 1.3 Data access

Simulation data that are associated with ongoing research projects are available to collaborators who have an account on the UW high performance computing cluster, `mox-hyak`<sup>16</sup>. Files are located in the `/gscratch/bumblereem/roms_for_public` folder and follow the naming scheme and Level 1–3 variable organization described above.

Simulation data associated with completed and published research projects are currently available on a PMEL-hosted THREDDS server<sup>17</sup>. To date, this includes the hindcast simulation and ACLIM Phase 1 CMIP5 simulations. Data on the THREDDS server is available in two different formats: individual files and aggregated datasets.

<sup>15</sup> <https://tinyurl.com/b10kvariablelist>

<sup>16</sup> see "Hyak: A Shared Scalable Compute Cluster for Research"

<sup>17</sup> THREDDS Server: <https://data.pmel.noaa.gov/aclim/thredds/>

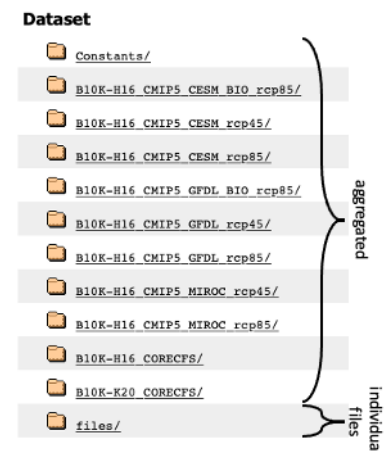


Figure 1: Current top-level file structure for Bering10K model output on the public THREDDS server.

The individual files (located under the "files" folder in the top-level catalog listing) follow the naming conventions described above, and each hold data associated with a single primary state variable over a 5-year block of the simulation. This file structure is very similar to the one used for collaborator-only data on the mox-hyak computer. These files can be accessed through either OPenDAP or HTTPS. We recommend using these file addresses for most users, especially those who wish to download local copies of the files (HTTPS recommended).

The aggregated files (listed by simulation name in the top-level catalog listing) combine all variables and all time steps from a simulation into a few access points (one for all Level 1 variables except constants, one for surface-averaged Level 2 variables, one for bottom-averaged Level 2 variables, etc.). The aggregated format may be more convenient to work with programmatically, since many variables can be accessed from a single url, but may be less user-friendly to those not familiar with ROMS output data (for example, state variables are not explicitly grouped with their associated coordinate variables). These endpoints support OPenDAP only. We recommend limiting data queries to manageable sizes to avoid timeouts and memory errors that may result when attempting to access large chunks of data.

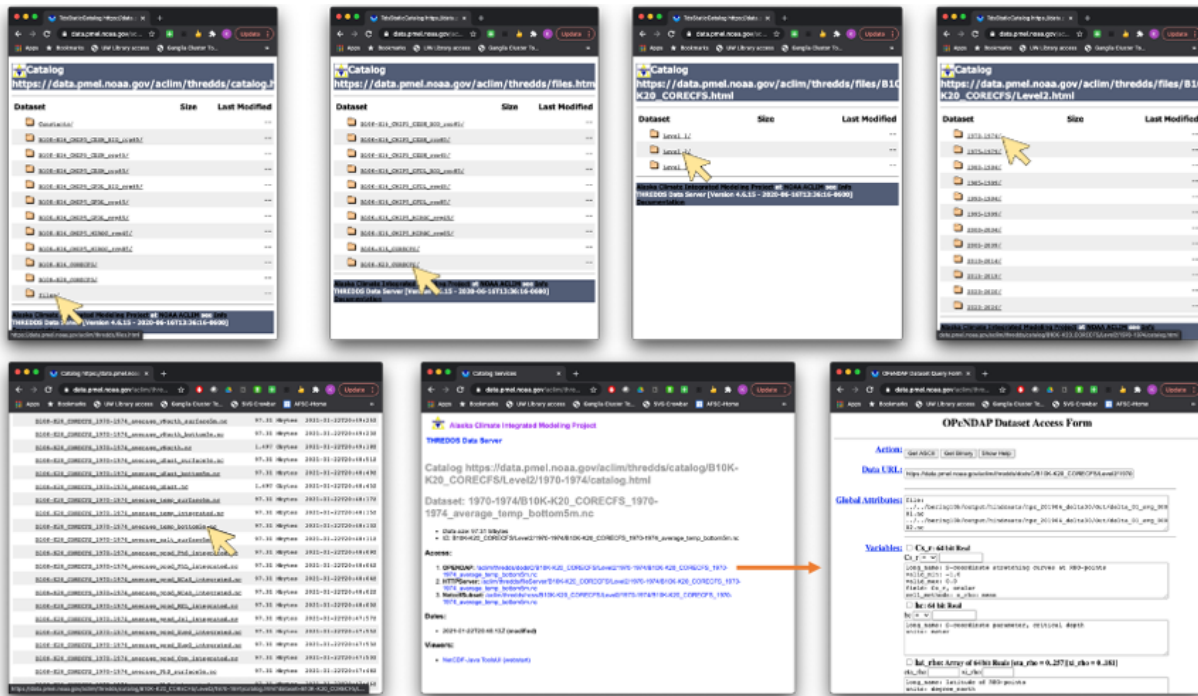


Figure 2: An example showing navigation through the THREDDS Data Server web interface. This example shows the steps needed to download a single file from the hindcast dataset (specifically, the 1970-1974 bottom temperature variable). Files can either be accessed via the OPenDAP url (lower right panel, link in the "Data URL" field), or via HTTPS (lower center panel, "HTTPS" link).

## 1.4 A few notes and cautions

### Stoichiometry in the BESTNPZ model

Some of the biological output variables (e.g. all the live critters) are saved using carbon-based units (mg C), while other (e.g. nutrients) are saved in nutrient-specific units (e.g. mmol N,  $\mu\text{mol Fe}$ ). Internally, the BESTNPZ model performs all of its calculations using nitrogen units, with a constant N:C:Fe ratio set by the input parameters  $x_i$  (N:C, completely unrelated to the  $x_i$ -coordinate) and FeC. In all of these simulations, Redfield-ish conversion factors were used (i.e. 106:16:0.0002 C:N:Fe, with conversion factors of 0.0126 mmolN/mgC and 0.0001667  $\mu\text{molFe/mgC}$ ).

### H16 vs K20 output variable differences

The H16 to K20 update to the BESTNPZ code included a number of bug fixes, a few changes to the underlying biological equations, an increase in vertical resolution, and also a complete overhaul of the internal structure of the biological source code to improve human readability (see Kearney et al., 2020 for more details).

The code overhaul included some adjustments to the netCDF variable attributes associated with the archived biological variables. Several of the variables were renamed (i.e. the `long_name` attribute in the K20 files vs H16 files may be different for the same state variable.) Except where noted, this is simply a semantic difference to better reflect the intended interpretation of each state variable. In some cases, this involved clarifying variable names that lagged behind prior code changes (for example, when detritus was split into slow- and fast-sinking state variables, or when microzooplankton size classes were merged into one). In others, we determined that the names could be misleading, and changed them accordingly (for example, the “Neocalanus” name implied a more species-specific focus than is actually present in the model equations, and was revised to “large-bodied copepods”; Ben was updated to clarify that it represents only benthic infauna, not epifauna or macrobenthos.)

There are a handful of places where things changed in a more meaningful way, and these changes are important to consider when working with the output data, especially if comparing between H16 and K20 simulations.

- Ice nutrient units: There was a bug in an input file (used for the H16 simulations, and later resurrected for some CMIP6 simulations) that led to the `IceN03` and `IceNH4` variables being labeled with units of  $\text{mg C m}^{-3}$ , while the values stored in those variables were actually in units of  $\text{mmol N m}^{-3}$ . While we attempted to fix this in most of the Level 1/2 output files, you may still find the incorrectly-labeled

units in some of the raw (Level 0) output. The  $\text{mmol N m}^{-3}$  units should be assumed for both H16 and K20 simulations.

- Ice algae units: In H16, IcePhL was archived using ice-layer-integrated units ( $\text{mg C m}^{-2}$ ). In K20, we shifted to reporting it in concentration ( $\text{mg C m}^{-3}$ ) to be consistent with other living state variables. The unit discrepancy here may be an input-related bug, similar to that seen in IceNO3 (i.e. the H16 version of the code may have actually been writing out volumetric values but the output was mislabeled); I have not been able to confirm this yet and so have not altered the label from the originals.
- Net production units: Internally, the BESTNPZ code calculates source/sink flux rates in layer-integrated units ( $\text{mmol N m}^{-2} \text{d}^{-1}$ ). In the H16 code, these were translated to volumetric rates ( $\text{mg C m}^{-3} \text{d}^{-1}$ ) for archiving. In the K20 code, we leave these in the layer-integrated units to be consistent with all the other flux rate diagnostic variables. Note that this difference is only present in the Level 1 3D depth-resolved net production variables; the Level 2 depth-integrated versions have all been translated to the same set of units.
- Large copepod and euphausiid production: In H16, the net secondary production for the two large copepod groups (NCaS and NCa0) and the two euphausiid groups (EupS and Eup0) were reported as single summed terms (`prod_NCa` and `prod_Eup`). In K20, this consolidation was removed, with production for each individual group reported. To compare between H16 and K20, users should replicate the summation (as well as the unit conversion described in the previous note).

## 2 The Bering10K coordinate system

### 2.1 The Bering10K grid

ROMS models, including the Bering10K model, use a curvilinear coordinate system in the horizontal grid. In the vertical, it uses a terrain-following s-coordinate.

The Bering10K horizontal grid includes relatively evenly-sized grid cells (approximately 10-km resolution), so unlike with some ROMS model grids, users don't need to worry about curving or distortion; the primary complicating factor is that it's rotated relative to geographic north-east coordinates. All output variables are defined in terms of the rotated  $\xi$  ( $x_i$ ) and  $\eta$  ( $\eta$ ) axes. From the point of view of the eastern shelf, the  $\xi$ -axis is approximately the cross-shelf direction, and the  $\eta$ -axis is approximately along-shelf.



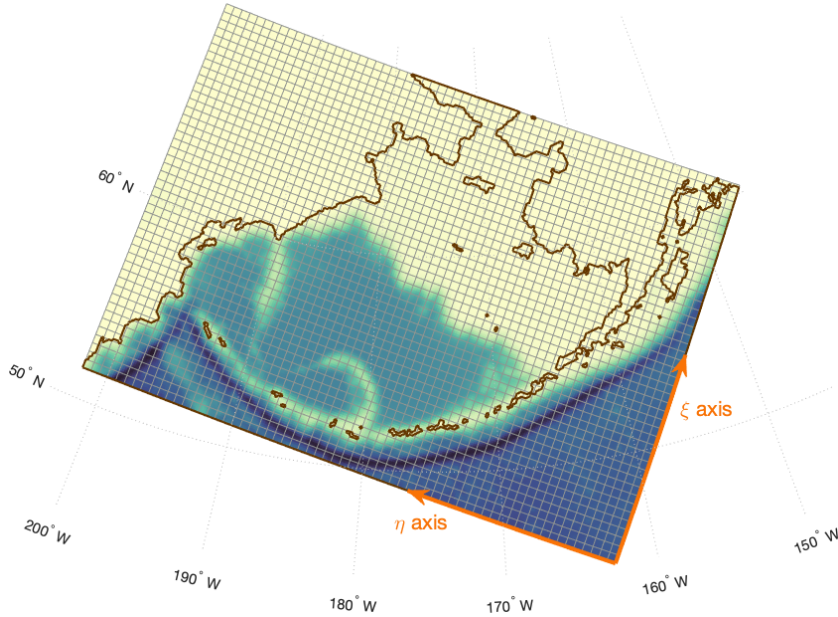


Figure 3: The Bering10K grid. Shading indicates model bathymetry (light to dark = shallow to deep). Note: Grid cells shown at 1/4 resolution for legibility.

In the output, the vertical coordinate is labeled as the  $s$  dimension. The depths of midpoints and edges of the 30 layers are a function of both the bottom depth and the free surface height. Because the free surface height changes over time (due to tides, winds, etc.), the exact depth of each layer also changes over time. These changes are generally small, but can become important in shallow water, where the changes in free surface height are significant relative to the total water depth. Depths can be calculated as follows:

$$z(n,k,j,i) = S(k,j,i) + \text{eta}(n,j,i) * (1 + S(k,j,i) / \text{depth}(j,i))$$

where

$$S(k,j,i) = \text{depth}_c * s(k) + (\text{depth}(j,i) - \text{depth}_c) * C(k)$$

The  $i$ ,  $j$ , and  $k$  indices correspond to the  $\zeta$ ,  $\eta$ , and  $s$  dimensions, respectively, and  $n$  is the time index. The  $s$ -coordinate variables ( $s\_rho$  and  $s\_w$ ) each include an attribute labeled "formula\_terms" that explains which ROMS output variables to substitute into these equations. When working with raw output files, the bottom depth values need to be retrieved from the grid file used to run the simulation; the remaining equation-term variables will be present in the output files themselves.

Post-processed output often includes pre-calculated values for the vertical depth grids among the Level 2 output (see below).

## 2.2 Output variable positions

ROMS uses a staggered Arakawa C-grid for its internal calculations. This means that certain properties, and the calculations involving those properties, are positioned in the center of grid cells, while others are positioned along the edges of the grid cells.

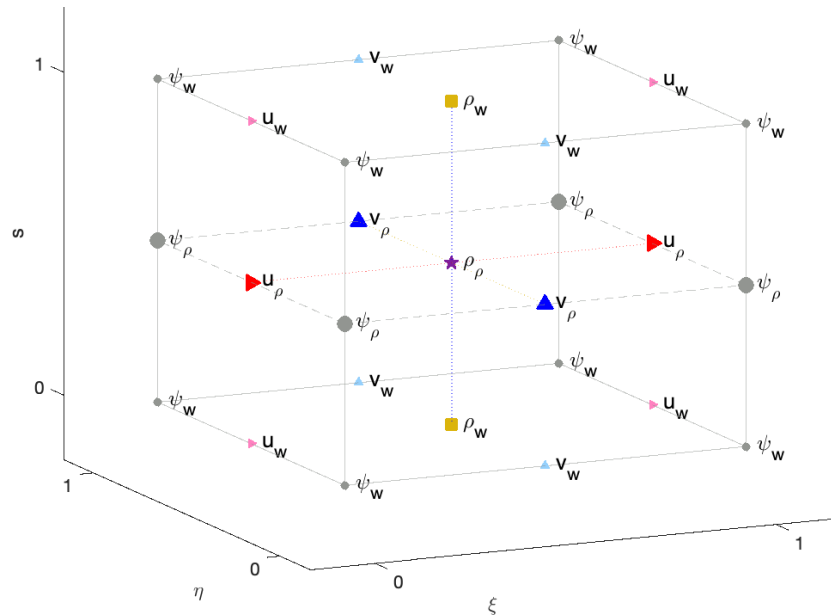


Figure 4: Staggered coordinate positions in the Bering10K grid.

In the vertical, there are two potential locations:

1. w-levels: on the edge of each layer
2.  $\rho$ - (rho-) level: near the center. The exact position of each  $\rho$  level follows the vertical coordinate equation given in the previous section, so it is *not* the center of the layer.

In the horizontal, there are 4 different grids:

1.  $\rho$ -grid: the center of each grid cell
2.  $\psi$ - (psi-) grid: the corners of each grid cell
3. u-grid: the midpoint of the right/left edges of each grid cell
4. v-grid: the midpoint of the top/bottom edges of each grid cell

2D output variables can be positioned on any of these four horizontal grids. In the variable list file, these positions are referred to as r2d, p2d, u2d, and v2d, respectively.

3D output can be positioned on one of 5 locations that combine the horizontal and vertical grids:

1.  $\rho$ -points: horizontal  $\rho$  and vertical  $\rho$  (purple star)
2.  $\psi$ -points: horizontal  $\psi$  and vertical  $\rho$  (large gray circle)
3. u-points: horizontal u and vertical  $\rho$  (red left triangle)
4. v-points: horizontal v and vertical  $\rho$  (blue up triangle)
5. w-points: horizontal  $\rho$  and vertical w (gold square)

These positions are referred to as r3d, p3d, u3d, v3d, and w3d in the variable list file.

While the psi-points represent the edges of the physical grid, the output files also hold an extra row/column of rho-points (along with the u/v-points in between); these are used within the ROMS calculations to hold the lateral boundary conditions. These extra boundary points can usually be ignored. The following figure shows how the indices of each grid align with each other at the corners of the domain (assuming 1-based indexing)

Latitude and longitude coordinates for the 4 horizontal grids can be found in the Bering\_grid\_withFeast.nc file; the appropriate lat/lon variables are also included in each output file for all Level 1-2 output.

### 2.3 Working with u and v

The u and v vectors in ROMS output are defined along the model's  $\xi$  and  $\eta$  axes, not along geographic east and north. To rotate between geographic (LON,LAT) and grid-aligned (XI,ETA) components, the following equations apply:

$$\begin{aligned}
 u(\text{LON}, \text{LAT}) &= u(\text{XI}, \text{ETA}) * \cos(\text{angle}(i, j)) - v(\text{XI}, \text{ETA}) * \sin(\text{angle}(i, j)) \\
 v(\text{LON}, \text{LAT}) &= v(\text{XI}, \text{ETA}) * \cos(\text{angle}(i, j)) + u(\text{XI}, \text{ETA}) * \sin(\text{angle}(i, j)) \\
 u(\text{XI}, \text{ETA}) &= u(\text{LON}, \text{LAT}) * \cos(\text{angle}(i, j)) + v(\text{LON}, \text{LAT}) * \sin(\text{angle}(i, j)) \\
 v(\text{XI}, \text{ETA}) &= v(\text{LON}, \text{LAT}) * \cos(\text{angle}(i, j)) - u(\text{LON}, \text{LAT}) * \sin(\text{angle}(i, j))
 \end{aligned}$$

Angles can be found in the Bering\_grid\_withFeast.nc file.

Also, because the u- and v-components are staggered on different grids, one often wants to co-locate these onto either the  $\rho$ - or  $\psi$ -grids to calculate things like speed and direction, or to plot vectors. Moving onto the  $\psi$ -grid is more consistent with the underlying physics, and will maintain proper land masking:

$$\begin{aligned}
 \text{Upsi}(i, j) &= 0.5 * ( U(i, j-1) + U(i, j) ) \\
 \text{Vpsi}(i, j) &= 0.5 * ( V(i-1, j) + V(i, j) )
 \end{aligned}$$

Moving onto the  $\rho$ -grid is often more useful for biological purposes, since all the of biological tracer variables, as well as temperature and salinity, are located on that grid.

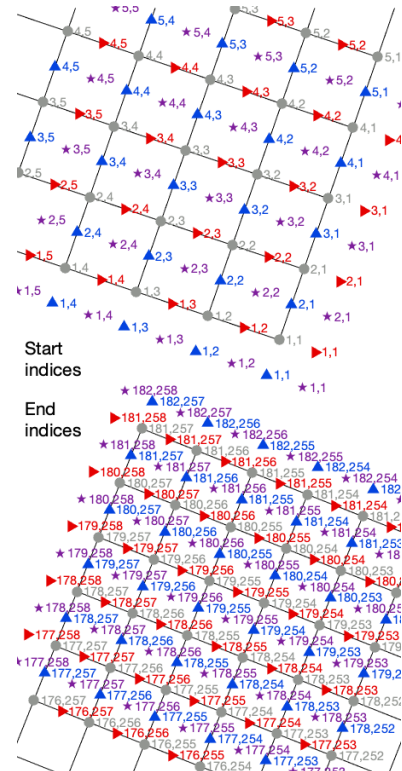


Figure 5: Indexing (one-based) at the corners of the Bering10K grid. The first number in each pair corresponds to the i ( $\xi$ ) index, and the second to the j ( $\eta$ ) index.

$$Urho(i,j) = 0.5 * ( U(i,j) + U(i+1,j) )$$

$$Vrho(i,j) = 0.5 * ( V(i,j) + V(i,j+1) )$$

## 2.4 Pre-calculated coordinate values

For ease of use, some Bering10K datasets are distributed with pre-calculated values that deal with the aforementioned coordinate complexities. When available, these will be found in the Level 2 output folder of a dataset.

- Pre-calculated depth values:  $z\_rho$ ,  $z\_w$ ,  $z\_u$ ,  $z\_v$ ,  $z\_psi$  variables correspond to  $\rho_\rho$ ,  $\rho_w$ ,  $u_\rho$ ,  $v_\rho$ , and  $\psi_\rho$  points, respectively, in the Fig. 4 above. Some datasets also include  $z\_psi\_w$ ,  $z\_u\_w$ , and  $z\_v\_w$ , which correspond to the  $\psi_w$ ,  $\psi_u$ , and  $\psi_v$  points in the above figure, respectively. No ROMS output variables are located on these last 3 grids, but these depths can be used to calculate layer thicknesses corresponding to 3D variables on the psi, u, or v grids, respectively.
- Geo-rotated u and v:  $uEast$  and  $vNorth$  variables hold u- and v-data shifted onto the  $\rho$ -grid and expressed in terms of geographic east and north components rather than  $\zeta$  and  $\eta$  components.

## 3 The ACLIM indices

In support of the Alaska Climate Integrated Modeling (ACLIM) project, a number of different biophysical index timeseries were calculated based on the Bering10K simulations. These indices, which are found under the Level 3 output for the relevant simulations, provide the primary means of linking the physical and lower trophic level dynamics simulated by the Bering10K long-term forecast simulations to the ACLIM suite of upper trophic level and socioeconomic models; see Hollowed et al. [2020]<sup>18</sup> for further details.

The specific Bering10K output variables included in this dataset represent the combined needs of the ACLIM ecosystem model end users. Two different types of ACLIM index variable datasets are currently available, which we describe below.

### 3.1 Regional average time series

The regional average indices (ACLIMregion\_\*.nc) include values spatially-averaged over a number of predefined polygon regions across the Bering Sea shelf.

The polygons are primarily based on the AFSC groundfish survey strata polygons. These polygons are plotted above in black, and are labeled in the regional average index files by their stratum number.

<sup>18</sup> A. B. Hollowed, 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. Integrated modeling to evaluate climate change impacts on coupled social-ecological systems in Alaska. *Frontiers in Marine Science*, 6(January):1–18, 2020. doi: 10.3389/fmars.2019.00775

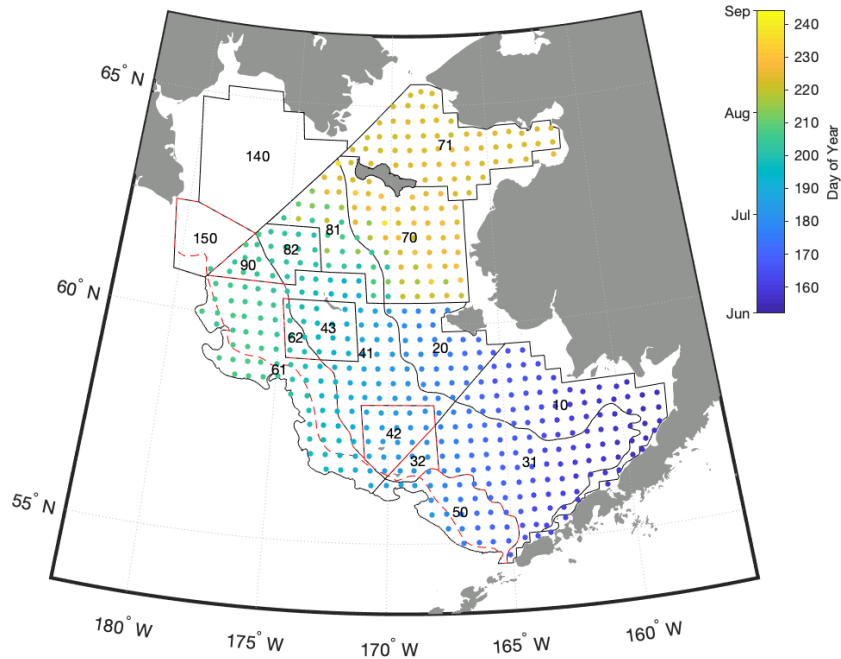


Figure 6: Regional polygons and survey-replication stations associated with the ACLIM indices.

The region files also include a modified set of polygons, identified by negative stratum numbers. These modified polygons, shown above with the red dashed lines, trim the outer polygons along the Bering10K model's 200-m isobath (which is displaced relative to the real-world 200-m isobath used to define the survey strata). When comparing real-world data to simulated data, we recommend using these modified polygon definitions to avoid mismatches stemming from the model-versus-real world bathymetry differences.

Regional indices are calculated using the same weekly time axis as in the underlying simulations. These can be further aggregated into annual or seasonal timeseries as needed. Note that the time axes of different simulations may not align perfectly with each other, due to differing start times and archiving parameters used for each simulation.

The region files include the strata number (`region`) and polygon area (`region_area`) for each polygon. See the `ACLIMregion.shp` shapefile for full geometry of the polygons.

### 3.2 Survey-replicated time series

The survey-replicated indices (`ACLIMsurveyrep_*.nc`) include values resampled at specific dates and times to mimic the annual groundfish survey sampling. In each year of the simulation, data is collected from a number of locations on the mean sample date for that location. See

Kearney et al. (in review) for further details of this survey-replication method.

The survey-replicated indices are annual timeseries. For ease of comparison, indices were extracted from all simulations using the same time and variable axes; this means that some variables may include placeholder fill values for years outside of the simulation time domain or for variables that were not archived for a particular simulation.

#### 4 Extended grid file regions

The geometry of a ROMS domain, including the location of the horizontal grid cells, rotation and curvature parameters associated with each grid cell, bathymetry values, etc. are contained within the grid file used for each simulation. The Bering10K grid file has also been supplemented with a number a regional subsetting datasets; these are not used by the ROMS code itself, but can be useful when analyzing the output of Bering10K simulations. These variables classify the rho-coordinate grid cells based on whether they fall within certain polygons. The following are currently present in the file:

- `surveystrata_updated`: Survey strata polygons associated with the annual Bering Sea continental shelf survey. This version was updated in 2019 to better characterize the northern Bering Sea; it does not include polygons outside of U.S. waters.
- `surveystrata_original`: Survey strata polygons associated with the annual Bering Sea continental shelf survey, pre-2019. This earlier version of the survey strata includes some slightly-shifted polygon boundaries relative to the newer version; it includes polygons that encompass the Gulf of Anadyr and western shelf as well as the eastern shelf.
- `surveystrata_comboeast`: This combination version of the survey strata uses the updated version of the eastern shelf polygons, supplemented by polygons from the older version that lie on the eastern shelf but in Russian waters. This is the version used for the ACLIM regional indices (see Fig. 6).
- `bsierp_marine_regions`: Marine regions developed under Bering Ecosystem Study (BEST) and Bering Sea Integrated Ecosystem Research Program (BSIERP) [Sigler et al., 2010] and documented in Ortiz and Greig [2013] (DOI: 10.5065/D6DF6P6C).
- `stations_regions`: These regions are less formally-defined than any of the above ones. They are usually used for model calibration

and validation purposes, and also correspond to the station coordinates in many Bering10K simulations (see `*sta.nc` output files). They consist of the Ortiz and Greig [2013] marine regions (1-16) plus 17: Gulf of Anadyr, 18: Karaginskiy Gulf, 19: Aleutian Basin, 20: Bowers Basin, 21: Komandor Basin, 22: Bowers Ridge, and 23: Shirshov Ridge. Regions 17-23 were defined using a combination of marine polygons from the Natural Earth dataset<sup>19</sup>, the IHO-IOC GEBCO Gazetteer of Undersea Feature Names<sup>20</sup>, and bathymetry contour lines.

- `shelf_mismatch`: Within the Bering10K bathymetry, the continental shelf break region has been smoothed<sup>21</sup>; as a result, the simulated shelf is narrower than the real-world shelf. This mask indicates where grid cells correspond to the model's slope or off-shelf regions but to the real-world on-shelf region.

<sup>19</sup> <https://www.naturalearthdata.com>

<sup>20</sup> [www.gebco.net](http://www.gebco.net)

<sup>21</sup> This avoids artificial horizontal pressure gradients in the numerical simulation. See Sikirić et al. [2009] for a summary of why  $\sigma$ -coordinate models like ROMS may have problems with steep topography.

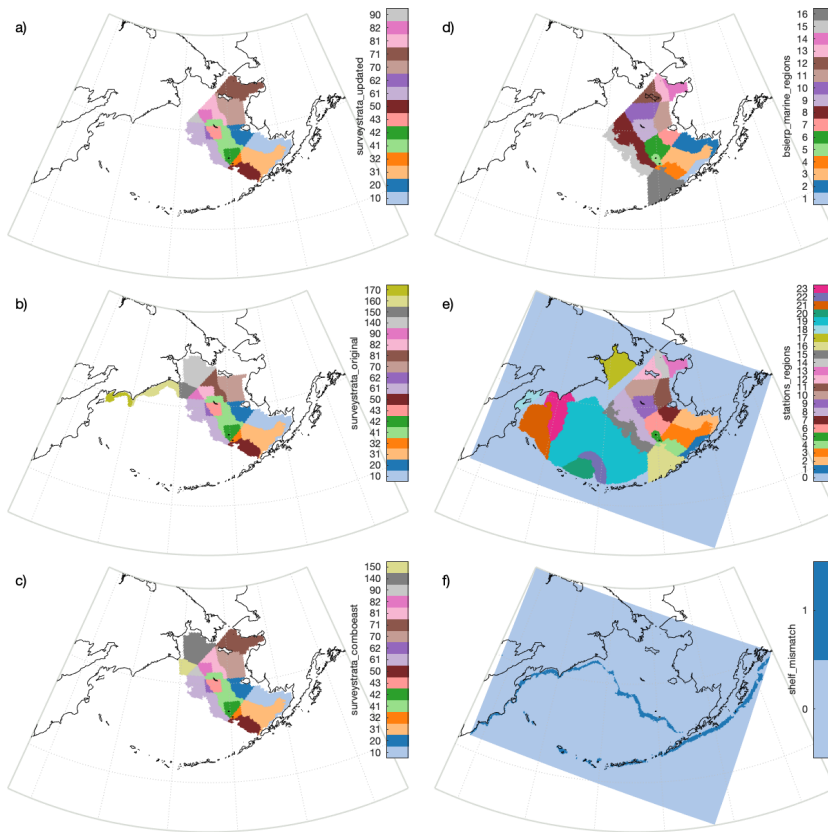


Figure 7: Maps of the extended grid file datasets, including a) Updated Bering Sea continental shelf survey strata (`surveystrata_updated`), b) Original survey strata (`surveystrata_original`), c) Combined eastern shelf survey strata (`surveystrata_comboeast`), d) BSIERP marine regions (`bsierp_marine_regions`), e) station-file analysis regions (`stations_regions`), and f) shelf mismatch regions (`shelf_mismatch`).

## References

- A. J. Hermann, E. N. Curchitser, K. Hedstrom, W. Cheng, N. A. Bond, M. Wang, K. Aydin, P. J. Stabeno, E. D. Cokelet, and G. A. Gibson. Projected future biophysical states of the Bering Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 134:30–47, 2016. ISSN 09670645. doi: 10.1016/j.dsr2.2015.11.001. URL <http://dx.doi.org/10.1016/j.dsr2.2015.11.001>.
- A. B. Hollowed, 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. Integrated modeling to evaluate climate change impacts on coupled social-ecological systems in Alaska. *Frontiers in Marine Science*, 6(January):1–18, 2020. doi: 10.3389/fmars.2019.00775.
- K. Kearney, A. Hermann, W. Cheng, I. Ortiz, and K. Aydin. 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, 2020. doi: 10.5194/gmd-13-597-2020. URL <https://www.geosci-model-dev.net/13/597/2020/>.
- W. G. Large and S. G. Yeager. The global climatology of an interannually varying air–sea flux data set. *Climate dynamics*, 33(2):341–364, 2009. ISSN 0930-7575.
- I. Ortiz and F. Greig. Marine Regions Boundary Data for the Bering Shelf and Slope, Version 1.0, 2013.
- D. J. Pilcher, D. M. Naiman, J. N. Cross, A. J. Hermann, S. A. Siedlecki, G. A. Gibson, and J. T. Mathis. Modeled Effect of Coastal Biogeochemical Processes, Climate Variability, and Ocean Acidification on Aragonite Saturation State in the Bering Sea. *Frontiers in Marine Science*, 5(January):1–18, 2019. doi: 10.3389/fmars.2018.00508.
- S. Saha, 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. The NCEP climate forecast system reanalysis. *Bulletin*



*of the American Meteorological Society*, 91(8):1015–1057, 2010. ISSN 00030007. doi: 10.1175/2010BAMS3001.1.

M. F. Sigler, H. R. Harvey, C. J. Ashjian, M. W. Lomas, J. M. Napp, P. J. Stabeno, and T. I. Van Pelt. How Does Climate Change Affect the Bering Sea Ecosystem? *Eos Trans. AGU*, 91(48):457–458, 2010.

M. D. Sikirić, I. Janeković, and M. Kuzmić. A new approach to bathymetry smoothing in sigma-coordinate ocean models. *Ocean Modelling*, 29(2):128–136, 1 2009. ISSN 14635003. doi: 10.1016/j.ocemod.2009.03.009. URL <http://linkinghub.elsevier.com/retrieve/pii/S1463500309000742>.