**Table 1.** Description of the fields needed to describe the creation of your dataset.

|  |  |
| --- | --- |
| **Title of dataset** | An absorption-based approach to improved estimates of phytoplankton biomass and net primary production |
| **URL of dataset** | [*https://seabass.gsfc.nasa.gov/*](https://seabass.gsfc.nasa.gov/) *(raw data)*  [*https://www.bodc.ac.uk/*](https://www.bodc.ac.uk/) *(raw data)*  [*https://github.com/jfox-osu*](https://github.com/jfox-osu) *(processed data)* |
| **Abstract** | *For this study field data from the* North Atlantic Aerosol and Marine Ecosystem Study and Export Processes in the Ocean from Remote Sensing programs and the 22nd Atlantic Meridional Transect were used. NAAMES and EXPORTS programs were funded by the NASA and all data used are in the SeaBASS data repository which is accessible to all. The AMT program is funded by NERC and is openly available in the UK oceanographic database BODC. Processed data files are also provided at the personal Github data repository of the corresponding author. |
| **Keywords** | *NAAMES, EXPORTS, AMT, net primary production, backscatter, optics* |
| **Lead author for the dataset** | *James Fox* |
| **Title and position of lead author** | *Postdoctoral scholar* |
| **Organization and address of lead author** | Department of Microbiology, Oregon State University, Corvallis, OR, United States. |
| **Email address of lead author** | *James.fox@oregonstate.edu* |
| **Additional authors or contributors to the dataset** | Sasha Kramer2, Jason Graff3, Michael J. Behrenfeld3,Emmanuel Boss4, Gavin Tilstone5, Kimberly H. Halsey1 |
| **Organization associated with the data** | 1.Department of Microbiology, Oregon State University, Corvallis, OR, United States.  2. Interdepartmental Graduate Program in Marine Science, University of California, Santa Barbara, Santa Barbara, CA, United States.  3. Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, United States.  4. School of Marine Sciences, University of Maine, Orono, ME, United States.  5. Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK |
| **Funding** | *List the principle investigator names, title of grant, funding agency, and funding identification number*  James Fox, Global assessment of climate driven trends in marine primary productivity, National Aeronautics and Space Administration (NASA), 80NSSC21K0421  Michael Behrenfeld:   * First steps - Linking remotely-detectable optical signals,   photic layer plankton properties, and export flux, NASA, 80NSSC17K568   * TheNorth Atlantic Aerosol and Marine Ecosystem Study, NASA, NNX15AAF30G   Kimberly Halsey, First steps - TheNorth Atlantic Aerosol and Marine Ecosystem Study, NASA. NNX15AE70G  Gavin Tilstone*,* Climate Linked Atlantic Sector Science, UK Natural Environment Research Council. NE/R015953/1 |
| **License** | [***CCO***](https://creativecommons.org/publicdomain/zero/1.0/) |
| **Geographic location – verbal description** | *Atlantic Ocean, Pacific Ocean* |
| **Geographic coverage bounding coordinates** | *Ocean Station Papa in sub-Arctic Pacific* |
| **Time frame - Begin date** | *06/28-2012* |
| **Time frame - End date** | *09/06/2018* |
| **General study design** | *Field observations* |
| **Methods description** | *Describe the steps followed in the study within the above study design. Please be specific, include instrument descriptions, or point to a protocol online, if this is a data compilation please specify datasets used, preferably their DOI or URL plus general citation information in the table at the end of this document.*  During each study, a WetLabs AC-S provided hyperspectral particulate attenuation and absorption measurements, while a WetLabs ECO-BB3 measured angular scatterance at three wavelengths (470, 532, and 660 nm). These measurements were collected in line using the ships underway water supply. Net primary production was measured using 14C radioisotopes during all cruises. A network-based community detection analysis was used to determine the dominant pigment-based taxon of the phytoplankton community following Kramer et al. (2020).  *Data files and protocol description for AC-S and ECO-BB3 available at:* [*https://seabass.gsfc.nasa.gov/archive/MAINE/boss/NAAMES/naames\_1/archive*](https://seabass.gsfc.nasa.gov/archive/MAINE/boss/NAAMES/naames_1/archive)  [*https://seabass.gsfc.nasa.gov/archive/MAINE/boss/EXPORTS/exportsnp/archive*](https://seabass.gsfc.nasa.gov/archive/MAINE/boss/EXPORTS/exportsnp/archive)  *Data files and protocol description for AC-S and ECO-BB3 available at:*  [*https://seabass.gsfc.nasa.gov/archive/OSU/NAAMES/archive/NPP*](https://seabass.gsfc.nasa.gov/archive/OSU/NAAMES/archive/NPP)  [*https://seabass.gsfc.nasa.gov/archive/OSU/behrenfeld/EXPORTS/EXPORTSNP/archive*](https://seabass.gsfc.nasa.gov/archive/OSU/behrenfeld/EXPORTS/EXPORTSNP/archive)  [*https://seabass.gsfc.nasa.gov/archive/PML/AMT/AMT22/archive*](https://seabass.gsfc.nasa.gov/archive/PML/AMT/AMT22/archive)  Kramer, S. J., Siegel, D. A. and Graff, J. R. (2020) ‘Phytoplankton Community Composition Determined From Co-variability Among Phytoplankton Pigments From the NAAMES Field Campaign’, *Frontiers in Marine Science*. Frontiers Media S.A., 7. |
| **Laboratory, field, or other analytical methods** | *Describe the lab, field, or other processing methods for each variable included in the data table. This section may, and should, be long. You should insert additional rows in this table to complete this section.* |
| **Taxonomic species or groups** | *diatoms, dinoflagellates, haptophytes, cyanobacteria* |
| **Quality control** | *All data products have undergone rigorous QC following published standards. For optical data protocols followed Boss et al. 2007, 2013*  Boss, E., Picheral, M., Leeuw, T., Chase, A., Karsenti, E., Gorsky, G., Taylor, L., Slade, W., Ras, J. and Claustre, H. (2013) ‘The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition’, *Methods in Oceanography*. Elsevier Ltd, 7, pp. 52–62.  Boss, E. S., Collier, R., Larson, G., Fennel, K. and Pegau, W. S. (2007) ‘Measurements of spectral optical properties and their relation to biogeochemical variables and processes in Crater Lake, Crater Lake National Park, OR’, *Hydrobiologia*, 574(1), pp. 149–159. |
| **Additional information** | *Any additional information that may help future users of the data not included in the above rows, or in the table below.* |
|  |  |

**Table 2.** Data dictionary: description of the variables (i.e., columns) in EACH dataset. You must provide sufficient detail for another user to understand and use the data. If there are 10 variables (i.e., columns) in the dataset, then there should be 10 rows in this table that describe each column. Be sure to include all relevant information for your dataset, including the unique identifiers for your dataset or system, dates, replicate numbers, latitude and longitude of sampling locations, etc.

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/amt\_npp.csv*

Dataset description: *Modeled and measured net primary production data for the AMT22 cruise*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| yd | Day of year |  |  | Day of year | NA |
| Ez\_NPP\_cbbp | Model estimates of depth integrated net primary production using bbp scaled with a fixed scaling factor to obtain a proxy for phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| Ez\_NPP\_cmod | Model estimates of depth integrated net primary production using a modeled estimate of phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| Ez\_NPP\_cvar | Model estimates of depth integrated net primary production using bbp scaled with a variable scaling factor to obtain a proxy for phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| 14C | Estimates of net primary production made using 14C radioisotope experiments | mg C m-2 d-1 |  | Daily average | NA |
| Cruise | The cruise the data were collected during |  |  |  |  |

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/naames\_npp.csv*

Dataset description: *Modeled and measured net primary production data for all NAAMES cruises*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| yd | Day of year |  |  | Day of year | NA |
| Ez\_NPP\_cbbp | Model estimates of depth integrated net primary production using bbp scaled with a fixed scaling factor to obtain a proxy for phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| Ez\_NPP\_cmod | Model estimates of depth integrated net primary production using a modeled estimate of phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| Ez\_NPP\_cvar | Model estimates of depth integrated net primary production using bbp scaled with a variable scaling factor to obtain a proxy for phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| 14C | Estimates of net primary production made using 14C radioisotope experiments | mg C m-2 d-1 |  | Daily average | NA |
| Cruise | The cruise the data were collected during |  |  |  |  |

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/exp\_npp.csv*

Dataset description: *Modeled and measured net primary production data for the EXPORTS cruise*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| yd | Day of year |  |  | Day of year | NA |
| Ez\_NPP\_cbbp | Model estimates of depth integrated net primary production using bbp scaled with a fixed scaling factor to obtain a proxy for phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| Ez\_NPP\_cmod | Model estimates of depth integrated net primary production using a modeled estimate of phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| Ez\_NPP\_cvar | Model estimates of depth integrated net primary production using bbp scaled with a variable scaling factor to obtain a proxy for phytoplankton biomass | mg C m-2 d-1 |  | Daily average | NA |
| 14C | Estimates of net primary production made using 14C radioisotope experiments | mg C m-2 d-1 |  | Daily average | NA |
| Cruise | The cruise the data were collected during |  |  |  |  |

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/AMT\_model\_data.csv*

Dataset description: *The processed data used to model NPP for the AMT cruise*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| wt | Sea surface temperature | oC |  |  | NA |
| bbp | Particulate backscatter coefficient | m-1 |  |  | NA |
| chl | chlorophyll | mg m-3 |  | Daily average | NA |
| par | Photosynthetically available radiation | E m-2 h-1 |  | Daily integrated average/daylength | NA |
| mld | Mixed layer depth | m |  | Daily average | NA |
| yd | Day of year |  |  |  | NA |
| sf | Scaling factor required for conversion of bbp to phytoplankton carbon | Unitless in this form |  |  | NA |
| Lat | latitude | Decimal degrees |  |  | NA |
| lon | Longitude | Decimal degrees |  |  | NA |
| Ap490 | Absorption at 490 nm | m-1 |  |  | NA |
| C\_mod | Modeled estimates of phytoplankton carbon | Mg m-3 |  |  | NA |
| Date | date |  |  | YYYY-MM-DD | NA |
| doy | Day of year |  |  |  | NA |
| Kd\_M07 | The light attenuation coefficient | m-1 |  |  | NA |
| Ez1 | The euphotic depth | m |  |  | NA |
| fe | Iron limitation component | unitless |  |  | NA |
|  |  |  |  |  |  |

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/EXPORTS\_model\_data.csv*

Dataset description: *The processed data used to model NPP for the EXPORTS cruise*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| wt | Sea surface temperature | oC |  |  | NA |
| bbp | Particulate backscatter coefficient | m-1 |  |  | NA |
| chl | chlorophyll | mg m-3 |  | Daily average | NA |
| par | Photosynthetically available radiation | E m-2 h-1 |  | Daily integrated average/daylength | NA |
| mld | Mixed layer depth | m |  | Daily average | NA |
| yd | Day of year |  |  |  | NA |
| sf | Scaling factor required for conversion of bbp to phytoplankton carbon | Unitless in this form |  |  | NA |
| Lat | latitude | Decimal degrees |  |  | NA |
| lon | Longitude | Decimal degrees |  |  | NA |
| Ap490 | Absorption at 490 nm | m-1 |  |  | NA |
| C\_mod | Modeled estimates of phytoplankton carbon | Mg m-3 |  |  | NA |
| Date | date |  |  | YYYY-MM-DD | NA |
| doy | Day of year |  |  |  | NA |
| Kd\_M07 | The light attenuation coefficient | m-1 |  |  | NA |
| Ez1 | The euphotic depth | m |  |  | NA |
| fe | Iron limitation component | unitless |  |  | NA |
|  |  |  |  |  |  |

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/NAAMES\_model\_data.csv*

Dataset description: *The processed data used to model NPP for all NAAMES cruises*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| wt | Sea surface temperature | oC |  | Daily average | NA |
| bbp | Particulate backscatter coefficient | m-1 |  | Daily average | NA |
| chl | chlorophyll | mg m-3 |  | Daily average | NA |
| par | Photosynthetically available radiation | E m-2 h-1 |  | Daily integrated average/daylength | NA |
| mld | Mixed layer depth | m |  | Daily average | NA |
| yd | Day of year |  |  |  | NA |
| sf | Scaling factor required for conversion of bbp to phytoplankton carbon | Unitless in this form |  |  | NA |
| Lat | latitude | Decimal degrees |  |  | NA |
| lon | Longitude | Decimal degrees |  |  | NA |
| Ap490 | Absorption at 490 nm | m-1 |  |  | NA |
| C\_mod | Modeled estimates of phytoplankton carbon | Mg m-3 |  |  | NA |
| Date | date |  |  | YYYY-MM-DD | NA |
| doy | Day of year |  |  |  | NA |
| Kd\_M07 | The light attenuation coefficient | m-1 |  |  | NA |
| Ez1 | The euphotic depth | m |  |  | NA |
| fe | Iron limitation component | unitless |  |  | NA |
|  |  |  |  |  |  |

Dataset filename: *https://github.com/jfox-osu/variable\_slope/blob/main/Cmod\_all\_cruises.csv*

Dataset description: Daily averages ofthe particulate backscatter coefficient and modeled estimates of carbon, along with pigment-based determination of the dominate phytoplankton taxa. Data were used for Figure 2 in manuscript.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column name** | **Description** | **Units** | **Code explanation** | **Data format** | **Missing data code** |
| *The name of the variable in the dataset; avoid special characters, dashes and spaces* | *A detailed description of the variable* | *Units the variable is measured in* | *If you use codes in your column, please explain each code, such as: LR = Little Rock Lake; A=sample; etc.* | *State exactly how the data are stored; for dates, state how it is formatted, including time zone, etc.* | *If data are missing, indicate how they are stored, such as NULL, NA, blank cell, etc.* |
| Cruise | The cruise the data were collected during |  |  |  | NA |
| doy | Day of year |  |  |  | NA |
| C\_mod\_mean | Modeled estimates of phytoplankton carbon | Mg m-3 |  | Daily mean | NA |
| bbp469\_mean | Particulate backscatter coefficient | m-1 |  | Daily average | NA |
| C\_mod\_sd | Modeled estimates of phytoplankton carbon | Mg m-3 |  | Daily standard deviation | NA |
| bbp469\_sd | Particulate backscatter coefficient | m-1 |  | Daily standard deviation | NA |
| taxa | pigment-based determination of the dominate phytoplankton taxa |  |  |  | NA |

**Table 3. Data provenance**

If you used data derived from other sources, provide the information here so future users know where the data came from.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset title** | **Dataset DOI or URL** | **Creator (name & email)** | **Contact (name & email)** |
|  |  |  |  |
|  |  |  |  |

**Scripts/code (software) –** *OPTIONAL*

It is recommended that you also provide your scripts along with your data, although it is not required at this time in our journal.

|  |  |  |
| --- | --- | --- |
| **File name** | **Description** | **Scripting language** |
| *Plotting\_script (available soon)* | *Scripts used to create all figures* | *R* |
| *Model\_script (available soon)* | *Scripts used to model NPP* | *R* |

**Notes and Comments:**

**Scripts have not yet been deposited in the corresponding authors Github but will be in the immediate future. This will include the model scripts required to estimate NPP and the files used to create all figures.**