Software to enable ocean discoveries: a case study with ICESat-2 and Argo

Kelsey M Bisson[1], Romina Piunno[2], Zachary Fair[3], Rachel Tilling[4], Alessandro Di Bella[5], Nicole Abib[6], and Jessica Scheick[7]

[1]Oregon State University
[2]Unknown
[3]Goddard Space Flight Center
[4]NASA Goddard Space Flight Center
[5]DTU Space, National Space Institute, Technical University of Denmark
[6]University of Oregon
[7]University of New Hampshire

December 14, 2023

Abstract

Increased anthropogenic stressors (e.g., warming, acidification, wildfires and other extreme events) present complex observational challenges for Earth science, and no one sensor can 'do it all.' While many remote sensing technologies are available at present, scientific disciplines are often trained to use only a specific subset, greatly limiting scientific advancements. Here we present open-source software ('icepyx') that lowers the barrier for entry for two remote platforms offering vertically-resolved information about the ocean's subsurface: ICESat-2 (Ice, Cloud, and land Elevation Satellite 2) and Argo floats. icepyx provides object-oriented code for querying and downloading ICESat-2 and Argo data within a single analysis workflow. icepyx natively handles ICESat-2 data access and read-in; here we introduce the Query, Unify, Explore SpatioTemporal (QUEST) module as a framework for adapting icepyx to easily access and ingest other datasets and present Argo data as the initial use case. Seamless retrieval of coincident data from ICESat-2 and Argo enables improved targeted and exploratory studies across the cryosphere and open ocean realms. We close with recommendations for future work, a discussion of the value of open science, relevance of our work to upcoming satellite missions, and an invitation to join our programming community.

Hosted file

`982049_0_supp_11683552_s5knwp.docx` available at https://authorea.com/users/527646/articles/693451-software-to-enable-ocean-discoveries-a-case-study-with-icesat-2-and-argo

Software to enable ocean discoveries: a case study with ICESat-2 and Argo

K.M Bisson[1,2,=], R. Piunno[3,=], Z. Fair[4], R. Tilling[4,5], A. Di Bella[6], N. Abib[7], and J. Scheick[8]

[1] Ocean Biology and Biogeochemistry Program, NASA Headquarters, Washington D.C.
[2] Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR USA
[3] Department of Physics, University of Toronto, Toronto, ON Canada
[4] NASA Goddard Space and Flight Center, Greenbelt, MD USA
[5] University of Maryland
[6] ESA ESRIN
[7] University of Oregon, Eugene, OR USA
[8] University of New Hampshire, NH USA

Key Points

- We present open source-software that allows easy access to ICESat-2 and Argo data
- This software enables observations of vertical profiles in the ocean
- Additional data streams are planned with community input

Abstract Increased anthropogenic stressors (e.g., warming, acidification, wildfires and other extreme events) present complex observational challenges for Earth science, and no one sensor can 'do it all.' While many remote sensing technologies are available at present, scientific disciplines are often trained to use only a specific subset, greatly limiting scientific advancements. Here we present open-source software ('icepyx') that lowers the barrier for entry for two remote platforms offering vertically-resolved information about the ocean's subsurface: ICESat-2 (Ice, Cloud, and land Elevation Satellite 2) and Argo floats. icepyx provides object-oriented code for querying and downloading ICESat-2 and Argo data within a single analysis workflow. icepyx natively handles ICESat-2 data access and read-in; here we introduce the Query, Unify, Explore SpatioTemporal (QUEST) module as a framework for adapting icepyx to easily access and ingest other datasets and present Argo data as the initial use case. Seamless retrieval of coincident data from ICESat-2 and Argo enables improved targeted and exploratory studies across the cryosphere and open ocean realms. We close with recommendations for future work, a discussion of the value of open science, relevance of our work to upcoming satellite missions, and an invitation to join our programming community.

Plain Language Summary Earth is changing rapidly due to human actions, and many different observations are needed to meet the challenges of the 21st century. Scientists are often trained to only use a particular subset of tools, but there are other relevant tools that go unused. Here we provide software (QUEST, within icepyx) to bring together two sensors that are commonly associated with two different communities. The ICESat-2 satellite was launched primarily to improve understanding of icy regions, and Argo floats were invented to overcome sampling gaps in the ocean. Both tools provide up-to-date information about the water column on a global scale. We wrote software in an open-source language (Python) to ease the access of using these complex tools and advance scientific discovery for all disciplines while also growing a community of users. By virtue of the software being open source, anyone can join the community and make contributions, including to incorporate data from other sources. Ultimately,

we hope to grow the community, enabling more scientific discoveries to support societal solutions.

1. Introduction and background

Advances in remote sensing technologies across different sectors of Earth science offer a tremendous opportunity to explore multiple observations over a shared place and time. However, the historical separation between disciplines, even within earth sciences, presents a substantial challenge for user access and science implementation. In some cases, scientists trained within a specific discipline may not even be aware of other relevant data products that are publicly available for use. Although the oceanography community has made substantial progress in understanding the marine system with ocean-dedicated satellites starting with the Coastal Zone Color Scanner (Antoine et al., 1996), these platforms are limited to sunlit, cloud-free conditions, and limited information about the water column can be obtained. More advancements are possible by combining technologies intended for ocean work with those that were initially created for different purposes, ideally enabled by an open-source framework (Hostetler et al, 2018). For example, the recent use of an atmospheric lidar sensor (the Cloud-Aerosol Lidar with Orthogonal Polarization, CALIOP, on Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations, CALIPSO) in ocean studies led to exciting discoveries, including observing diel vertically migrating zooplankton from space (Behrenfeld et al., 2019), as well as advances in understanding that will improve conventional technologies (i.e., a seasonal bias in NASA ocean color data, Bisson et al., 2021a, 2023).

Two such technologies that have apparently different purposes yet strong compatibilities and applications in oceanography are Argo floats (Argo and biogeochemical (BGC), Argo, 2000) and the Ice Cloud, and land Elevation Satellite-2 (ICESat-2, Markus et al., 2017) mission launched in 2018. Argo floats move vertically within the water column as they drift in a semi-lagrangian manner with the currents, primarily offering vertical profiles of temperature, salinity, and depth. An additional group of Argo floats (BGC) are equipped with biological sensors to measure chlorophyll concentration, nitrate, oxygen, the particulate backscattering coefficient at 700 nm, and in some cases, photosynthetically active radiation (PAR) and spectral downwelling irradiance. A portion of Argo floats will transit underneath sea ice (Hague and Vichi, 2021), capturing precious data not possible using conventional methods, especially from satellites. The invention and deployment of Argo floats have transformed our understanding of ocean physics and biology, especially during times and in places inaccessible from ocean color satellite observations (e.g., wildfires, Tang et al., 2021; under-ice blooms, Horvat et al., 2022). Until recently, Argo floats were the only tool available to capture the structure of the upper water column on global scales.

In 2018, ICESat-2 was launched with a primary goal of studying polar regions, but it has technical capabilities to observe terrestrial and ocean ecosystems worldwide. The Advanced Topographic Laser Altimeter System (ATLAS), the sole instrument onboard ICESat-2 and the only powerful photon-counting lidar altimeter in orbit, contains a 532 nm laser with a pulse repetition rate of 10 kHz with three pairs of beams on the ground (Markus et al., 2017). Over the ocean, ICESat-2 data have been used to generate vertical profiles of particulate backscattering and light attenuation (Lu et al., 2020, 2021), map bathymetry accurate to ~0.5m (Parrish et al., 2019), and extract sea ice thickness data for contextualizing under ice phytoplankton phenology (Bisson and Cael, 2021), the latter of which was accomplished using software described herein.

With CALIPSO decommissioned, the only lidar satellite available for ocean observations presently flying is ICESat-2 (Behrenfeld et al., 2023).

Compared to ocean color data, which are typically manageable in aggregate because they are of relatively small size (available at https://oceancolor.gsfc.nasa.gov; note each day of data is ~1 GB or less), photon data from ICESat-2 are complex, large datasets that cannot be easily downloaded and stored for local use without substantial manipulation (e.g., just one day of ICESat-2's ATL03 data product is ~300 GB). While other missions (e.g., MODIS-Aqua, CALIPSO) have higher-level ocean data products, ICESat-2 ocean products are in development, and many ocean applications of ICESat-2 rely on lower-level ATL03 photon cloud data to derive subsurface optical information about the water column. Without open-source, collaborative software programs and community resource sharing, a substantial amount of prior knowledge is needed in order to appropriately and efficiently access and use ICESat-2 data for ocean applications. None of the recent studies using CALIOP for ocean particle and biology studies have made their code openly available or in an open-source language, limiting the degree to which satellite lidar analyses in the ocean can be reproduced and proliferated for different needs (Behrenfeld et al., 2013, 2017, 2019, 2022, Lu et al 2020, 2021, Lacour et al., 2020, Bisson and Cael, 2021, Bisson et al., 2021a,b, 2023).

To enable novel studies of coupled ICESat-2 and ocean data, we introduce an open-source Python module QUEST, housed and packaged within the icepyx library, a community and Python software library that simplifies the process of 'querying, obtaining, analyzing, and manipulating ICESat-2 datasets to enable scientific discovery' (Scheick et al., 2019, 2023). The module includes testing, documentation, and a tutorial for accessing coincident Argo and ICESat-2 data using QUEST. Our goal is to lower the access barrier to combining multiple datasets to advance our understanding of ocean/sea ice processes from polar to global scales (Figure 1). Here, we discuss the cultural and scientific value of collaborative approaches to working across disciplines and sensors (§2), including best practices for writing open source code, based on the authors' experiences in developing this workflow. Our software (§3) is object-oriented and written with flexibility so future datasets of interest can be included with ease (§4) for improved scientific application. We close with recommendations for future software capabilities, and we invite those interested to join our community and participate in ongoing efforts.

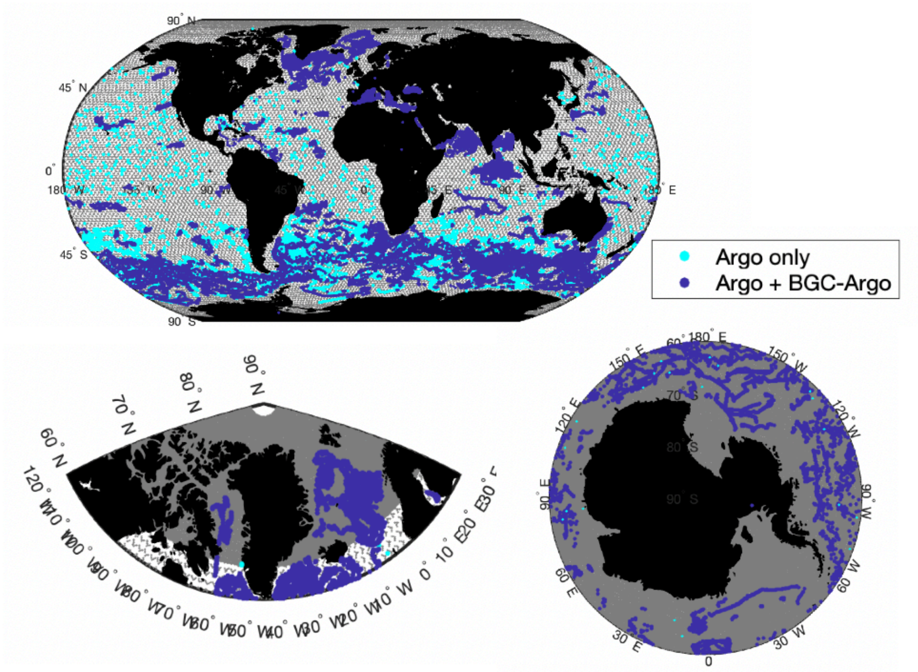

Figure 1. Location of ICEsat-2 reference ground tracks over the ocean (grey lines), Argo floats equipped with physical (i.e., temperature, salinity) sensors (cyan), and Argo floats with both biogeochemical and physical sensors (BGC, dark blue) globally and over both poles. ICESat-2 produces data from September 2018 to present, Argo (physical parameters) from 1999 to present, and BGC-Argo from 2016 to present.

2. Approach to programming and teamwork

2.1 Open science

The United States White House Office of Science and Technology Policy (OSTP) and National Science and Technology Council (NSTC) formally define "open science" as *"The principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity."* (U.S. OSTP and NSTC, 2023). Released in January 2023, this federal definition coincides with the recognition of 2023 as the Year of Open Science, a concept galvanized and promoted through NASA's Transform to Open Science (TOPS) Mission (https://science.nasa.gov/open-science/transform-to-open-science) which is part of the agency's broader Open-Source Science Initiative (OSSI) (https://science.nasa.gov/open-science-overview). While necessarily broad, this definition highlights the overarching principles that lead to many practices long ago adopted by many communities, and the open-source software community specifically. Here we highlight one of the many scientific achievements enabled by this type of trans-disciplinary, cross-platform, collaborative approach and hope to persuade readers to learn about and adopt relevant open science practices in their own teams and workflows. Our motivation stems from wanting to enable more cross-disciplinary discoveries through open science, in part because some of our previous work was not open. Developing QUEST provided a space to learn and exercise open science practices.

2.2 Our team

Our team met at the University of Washington's ICESat-2 2020 Virtual Hackweek (Arendt et al., 2020; Huppenkothen et al., 2018). During this event, project teams formed to collaborate on a pressing technical or research challenge. We identified a growing gap between ocean and cryosphere studies, namely the lack of ease with which one could download ICESat-2 data simultaneously with other data products of interest. We quickly created a proof of concept for combining Argo and ICESat-2 data. With no previous collaborations and hailing from different academic cultures, disciplines (ocean biology, glaciology, sea ice physics, physics), and time zones (from Pacific Time to Central European Standard Time), working together as a project group during the week-long event catalyzed a practice of virtual collaboration and support. At the time, only a few team members had experience working in Python and/or using version control tools (e.g. git, GitHub) to write code collaboratively. After the hackweek, the group continued meeting to create what ultimately became the QUEST module presented herein. To minimize the burden on already full schedules, we intentionally met for only an hour weekly, setting appropriately rigorous benchmarks for success while performing most of the work during these meetings. Importantly, during these supportive co-working online sessions, we not only wrote software but engaged in coding and collaboration best practices, building skills and learning from one another. In this way we created a culture of trust and transparency that enabled us to share our skill sets, make research and personal progress, and address challenges in real time.

2.3 Object-oriented development overview

A first step of our team's collaborative work was reformatting existing code to fully leverage the benefits Python's object-oriented structure has to offer. Details of changes to icepyx's architecture are outlined in §3.3. Object-oriented programming (OOP) is a common implementation feature of many popular open-source languages, including Python and JAVA, that organizes code through an object-centric perspective. An object is any "entity" possessing unique attributes and behaviors (Supplementary Figure 1). For example, a "person" object might include "name," "age," and "eye color" attributes and "sleep," "eat," and "express joy" behaviors. In the context of oceanography, a "water column" object would have "temperature," "salinity," and "chlorophyll," among other attributes. Structuring code in this object-oriented way has several benefits, most notable of which is modularity. Independent segments of code can be written simultaneously ("orthogonality" as defined by Thomas and Hunt, 2019) and then brought together like building blocks that interlink. This modularity enables multiple developers to individually write code segments independently and combine them later so long as there is an agreed upon input and output format ("Design by contract," Thomas and Hunt, 2019) between them. The modularity of OOP is conducive to easy maintenance because if one code segment needs to be modified, changes can be made without also propagating revisions through other segments, as long as the input/output criteria are met (see §3.3).

3. icepyx and QUEST: open-source software for ICESat-2 and Argo

3.1 What is icepyx?

icepyx is an open-source Python software package and community designed to enable collaboration and work with the large and complex data products from ICESat-2. icepyx was created at the first ICESat-2 Hackweek held in June 2019, less than a year after the launch of the satellite. During that event, data access methods were presented ad hoc, with new users required to carefully format tens of lines of code to submit a valid data access request or manually download individual files through a web browser. Since then, the package's capabilities have expanded as more users contribute their work. It now provides data access via download or in the cloud, visualization, and read-in capabilities. Critically, the software package provides a citable, tested, shared development framework that is publicly available and easily installable via GitHub (https://github.com/), PyPI (the Python Package Index; PyPI, 2023), and Conda (Anaconda, 2023), while the community provides a safe, supportive, communal learning space to build the skills required to effectively collaborate on code.

3.2 Specific software functionality

The entire process of querying and downloading (or accessing in the cloud) ICESat-2 data can be achieved with icepyx in three steps: (i) initialize the search with the 'Query' class, (ii) log into NASA Earthdata, and (iii) call the download functionality (or begin cloud reading). Below, we describe a few key programmatic features with which the user can interact. We encourage potential users to explore the icepyx documentation (https://icepyx.readthedocs.io/en/latest/) and examples (e.g., https://icepyx.readthedocs.io/en/latest/example_notebooks/IS2_data_access.html) to explore the full range of functionality available within the software. Users need a free Earthdata account (https://www.earthdata.nasa.gov) to download any ICEsat-2 data from the National Snow and Ice Data Center Distributed Active Archive Center (NSIDC-DAAC) or access it in the cloud; icepyx provides multiple authentication options for an individual to enter their credentials, including an in-notebook login. These options are showcased in the documentation and are not further addressed here.

The Query data object within icepyx allows the user to define their study parameters. Input variables include a string for the ICESat-2 product of interest (e.g., 'ATL03,' 'ATL07'), a spatial extent that can be represented as a bounding box or polygon (coordinates or geospatial polygon file), and a time window. A maximum of one spatial bounding box or search polygon is allowable per Query object instance, a limitation imposed by the data archive center but easily addressed with multiple Query objects. Additional search filters can be added for ICESat-2 queries if the user wishes to search for a specific product version, cycle, or reference ground track. The user can generate a map of their search region and view summary information about the data product using the methods available on the Query object (https://icepyx.readthedocs.io/en/latest/example_notebooks/IS2_data_access.html).

Configuration parameters required to search for and access data products are automatically generated by the software. The user can manually create, view, and update these parameters, but it is not required. After creating a Query object, the user can view the search results and metadata (e.g. avail_granules()). During data ordering and downloading, the user can additionally subset the file for specific parameters of interest (Supplementary Material) and supply options to change the file type (e.g., HDF5 to NetCDF4-CF (see show_custom_options() in the Query object for more details on available subsetting options).

3.3 QUEST

Here we present the Query Unify Explore SpatioTemporal (QUEST) module, which is an expansion of the icepyx Query class (§3.2). From the original icepyx Query object implementation, we modified the architecture to create a super class object called GenQuery. Parameters not specific to ICESat-2, such as spatial and temporal information, were isolated to be handled instead by GenQuery, making this information directly accessible to the QUEST module independent of the ICESat-2 Query functionality. In turn, QUEST uses this super class GenQuery to handle spatial and temporal data while also housing basic properties and functionalities common to all datasets (such as preparing data for plotting). These underlying changes are invisible to the user and take advantage of OOP's ability for high-level organization.

The QUEST module is designed to easily query, download, and perform simple operations on datasets complimentary to and including ICESat-2. Users specify spatiotemporal bounds for their investigation through creating a QUEST object. The user then utilizes this higher-level framework to call on subsets of the framework defined specifically for each type of dataset, providing any additional parameters important for obtaining or manipulating their dataset of interest (e.g., variables of interest). Attributes and behaviors that are common to all datasets and required by this higher-level framework are indicated in a template-like Dataset class and its per-dataset subclasses, with which the user is not intended to interact directly. This hierarchal system defines a structure for future developers to add functionality for additional datasets (§4, Supplementary Figure 1).

3.4 QUEST use case: Argo

Argo data are available for physical (pressure, temperature, salinity) and biogeochemical (chlorophyll-a, nitrate, dissolved oxygen, particulate backscatter, downwelling irradiance) parameters and in a range of data modes (i.e., real-time vs delayed). Real-time data are not quality controlled whereas delayed-mode data usually are, although some variates, including the particulate backscattering coefficient (b_{bp}), are not strictly quality controlled and need to be further examined by the user. Argo floats are numerous and the full dataset can be downloaded from two Global Data Assembly Centers (GDAC); GDAC data access does not permit a user to search and download for particular floats of interest unless the specific float number is known a priori (https://biogeochemical-argo.org/data-access.php). Downloading the entire Argo dataset is not feasible for users working locally on their computers due to size constraints, and working with numerous individual files is less efficient than working within a merged dataframe. Recently, Tucker et al. (2020) developed an application program interface (API) to query and download Argo data programmatically based on space/time windows through their web interface, Argovis. Here, we utilize the Argovis API within QUEST to query, download, and format delayed-mode Argo data of interest with minimum effort from the user. In this way, the user does not need to download Argo separate from their ICESat-2 Query, nor does the user need to download a static dataset from the GDAC. We present an example use case in the North Pacific, where ICESat-2 and Argo data are available < 5 days apart (Figure 2). In this case, the depth information available from ICESat-2 appears representative of the rough mixed layer depth (given by the temperature profile). While ICESat-2 has been used to generate vertical profiles in the ocean, it is not clear that these signals can be wholly attributed to phytoplankton, because particles, bubbles, and surface glint also have a role. By coupling nearby Argo observations with ICESat-2 data, one can more rigorously assess both datasets in tandem, improving the use of ICESat-2 to address ocean biology and biogeochemistry (Table 1). In the future, it may be possible to assess stratification in the upper water column from ICESat-2 photon clouds, but ancillary data such as Argo are needed to facilitate these comparisons and identify uncertainties.

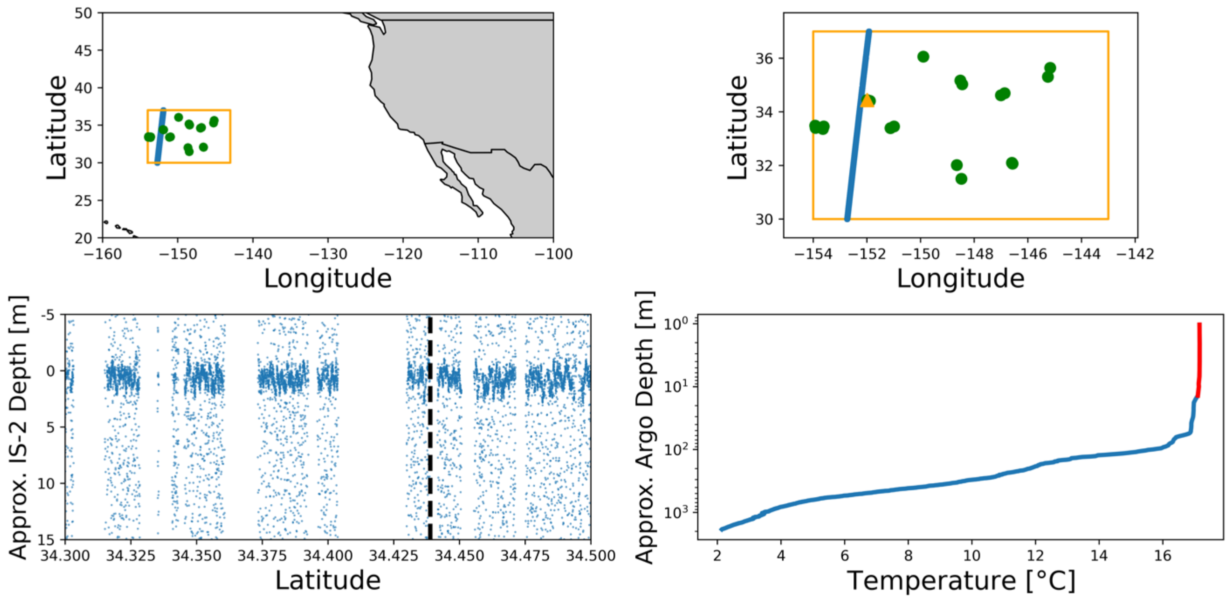

Figure 2. (Top left) Map of ICESat-2 (blue) and Argo (green) data within the icepyx bounding box (yellow). (Top right) Zoomed in view of spatial area with closest Argo profile selected in the yellow triangle. (Bottom left) Height versus latitude of ICESat-2 photons in the subsurface, with Argo location (black dashed line). (Bottom right). Depth versus temperature from Argo profile, with ICESat-2 vertical extent highlighted in red.

4. Steps and scientific value of adding a new dataset to icepyx

We anticipate many current and future datasets can be included within QUEST to greatly amplify the opportunity for scientific discovery at the nexus of disciplines. For example, PACE (the Plankton, Aerosols, Clouds, Ocean Ecosystems) mission (Werdell et al. 2019) will supply hyperspectral and polarized data on global scales, and GLIMR (Geosynchronous Littoral Imaging and Monitoring Radiometer, Salisbury, 2022) will provide hourly data and thereby increase the likelihood of synergies with ICESat-2 in the Gulf of Mexico region, where GLIMR is targeted to observe. Future progress may be enabled by cloud computing and subsetting procedures (described herein) that minimize the computational fluency required by the user to access data. We designed the QUEST module of icepyx to leverage object-oriented strengths in large part so that additional datasets can easily be added so long as an API is available, and a roadmap for adding a new dataset is provided in Figure 3.

The software we describe herein can be used to facilitate a number of targeted and exploratory studies across the interface of cryosphere and ocean studies (Table 1). For example, Argo profiles provide temperature and salinity that can be used to contextualize and test suspected sea ice melting events, and the optical sensors on Argo floats can be used to quantify glacial silt in tandem with ICESat-2 measurements of glacial activity. Argo floats also provide an important link between sea ice physics and ocean biology, which is needed given the rapid rate of Arctic warming (Rantanen et al., 2022) and associated biological changes (Ardyna and Arrigo, 2020). ICESat-2 and Argo data could also be used in tandem to generate vertical profiles of light attenuation or particulate backscattering in the ocean, which would enable fine scale exploratory studies in the upper-ocean. Fundamentally, icepyx facilitates direct comparison with ICESat-2 and Argo observations, which will only become more plentiful in the coming years.

Table 1. List of ICESat-2 products relevant for ocean studies

Product	What is it?	Used for?	References (Ocean focus)
ATL03	Global geolocated photon data	• Deriving optical information (light attenuation coefficient, b_{bp}) in coastal & global waters • Bathymetry in shallow waters	Lu et al. 2020, 2021 Eidam et al, 2022, 2023 Parrish et al, 2019
ATL07 (ATL20)	Polar sea ice elevation (Gridded sea ice freeboard)	• Sea ice freeboard • Sea ice lead identification	Bisson and Cael, 2021 Horvat et al, 2022
ATL12 (ATL19)	Ocean Elevation (Gridded sea surface height)	• Sea surface height	Bagnardi et al, 2021

Figure 3. Workflow illustrating steps of adding a new dataset to QUEST module.

5. A note on best practices

Python is one of the more forgiving languages in which to code, and this increased flexibility reduces both learning barriers and development time; however, sometimes more rigidly formatted code can benefit a project. Access modifiers are one such element that offer more rigidity but are not formally supported in the Python language. As the name suggests, access modifiers are syntax which modify access to objects. An object may be public, protected, or private. Public objects may be accessed anywhere in the program. Protected objects are only accessible within a class and its subclasses. As an example within the context of icepyx, "Dataset" is a high-level object from which more specific datasets extend including "Argo" and "ICESat-2". An attribute common to all datasets is the geographical location at which the data were collected. The region of interest is an attribute that the user will specify, regardless of the dataset being queried. API calls require geographic boundaries to be formatted in a specific way. Requiring the user to manually reformat the geographical region for each API call would be both tedious for the user and leave unnecessary room for error. This reformatting is best done on the backend via a function _fmt_coordinates(), with which the user should never interact. It is therefore best practice to designate this function as protected. That is to say, the higher-level "parent class", Dataset, possesses a generic _fmt_coordinates() function that is inherited by its "child classes", ICESat-2 and Argo. The specific child classes have access to the generic functionality, though the developer may also override _fmt_coordinates() within the child class itself to cater the formatting to the API being called. The take-away from this is that _fmt_coordinates() can be inherited by children of a class, but should not be called outside of the (sub)class itself.

The most restrictive access modifier is "private." This prohibits access to an object outside of a class. There are no private variables in icepyx at this time, in part because objects in Python are public by default and there is no true way of restricting access to objects. Access modifiers are built on a type of "honor system" in which the programmer is expected to respect access recommendations. Protected objects are prefixed with a single underscore, and private objects are prefixed with a double underscore. The end user is expected not to interact with objects with private or protected designations.

The best practices used in the development of icepyx extend beyond those visible in the code. Test Driven Development (TDD) is a school of software development whereby the program is written in response to test cases. This process begins with establishing the desired functionality, writing test cases to reflect that functionality, and finally writing code to achieve that functionality. Test cases are often thought to simply verify the program is behaving as expected; however, TDD encourages the developer to consider how the end product will be used. "Design by contract" and "orthogonality" are among the recommendations presented by Thomas & Hunt (2019) used explicitly in icepyx's QUEST module. The term "Orthogonality" signifies segments of code which are independent of one another. That is to say the inner workings of one segment should not affect the behavior of another segment. "Design by contract" offers a framework though which orthogonal code segments may interact. The developer decides on a contract of preconditions and post conditions to which the program should adhere. In the context of QUEST, ICESat-2 and Argo objects are independent of one another. There is, however, an agreed upon contract established by the higher-level "Dataset" class that exists solely on the backend which enforces input and output types expected by each of the two specific datasets.

6. Summary, and the value of open-source science to facilitate cross-disciplinary collaborations

Here we have introduced and described our efforts to build the QUEST module within icepyx, including architectural modifications to meet software development best practices and provide a superclass structure to readily accommodate future geophysical datasets. We have illustrated the science possibilities enabled by QUEST by incorporating physical and biogeochemical Argo data with ICESat-2 tracks as a case study. Future advancements will come by adding other datasets to QUEST and expanding upon this initial exploration of coincident data. The science community needs to embrace the philosophy that integrating technologies is required for ground-breaking advances, not only to achieve closure in the measured parameter of interest, but also to greatly extend what's possible from any one sensor alone. ICESat-2 and Argo are the only platforms that offer near real-time, global scale, vertically-resolved subsurface information about ocean biology and biogeochemistry at present; future missions will be easily included through our creation of shared, open computational pipelines and infrastructure.

Open-source science (OSS) is a powerful concept offering free and unlimited data access, fully documented open software and algorithms, fully transparent processes and reproducibility, and a teaching culture (https://www.earthdata.nasa.gov/esds/open-science). OSS and its adoption catalyzes cross-disciplinary conversations surrounding best practices for collaboration, ultimately enhancing community and scientific rigor. Proprietary software, lack of code sharing, and ambiguous methodologies hurt our potential for meaningful collaborations. As more technologies are developed and innovated, the need for transparency and data sharing will only grow. As we have described here, the QUEST module within icepyx provides a generalized framework such that future studies incorporating multiple sensors are not only possible, but could become routine and accessible even for novice developers.

Acknowledgements We are grateful to the icepyx community for their support as well as the organizing committee of the University of Washington ICESat-2 hackweek. KB acknowledges NASA grant # 80NSSC20K0970, JS acknowledges NASA grant #80NSSC21K0505. ZF thanks #NPP168273S under the NASA Postdoctoral Program.

Open Research Our software is freely and openly available at https://github.com/icesat2py/icepyx, and was used to download ICESat-2 and Argo data in this use case. Data from ICESat-2 used in this study are freely available at https://nsidc.org/home and are accessed through our software program, icepyx (Scheick et al., 2019, 2023), Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (http://doi.org/10.17882/42182). The Argo Program is part of the Global Ocean Observing System.

References

Anaconda (2023). *Anaconda Software Distribution*. Computer software. Vers. 2-2.7.3. Anaconda, Feb. 2023. <https://anaconda.com>.

Antoine, D., André, J. M., & Morel, A. (1996). Oceanic primary production: 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. *Global biogeochemical cycles*, *10*(1), 57-69.

Argo (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182

Ardyna, M., & Arrigo, K. R. (2020). Phytoplankton dynamics in a changing Arctic Ocean. *Nature Climate Change*, *10*(10), 892-903.

Arendt, Anthony, Scheick, Jessica, Shean, David, Buckley, Ellen, Grigsby, Shane, Haley, Charley, Heagy, Lindsey, Mohajerani, Yara, Neumann, Tom, Nilsson, Johan, Markus, Thorsten, Paolo, Fernando S., Perez, Fernando, Petty, Alek, Schweiger, Axel, Smith, Benjamin, Steiker, Amy, Alvis, Sebastian, Henderson, Scott, … Sutterley, Tyler. (2020). 2020 ICESat-2 Hackweek Tutorials (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.3966463

Behrenfeld, M. J., Lorenzoni, L., Hu, Y., Bisson, K. M., Hostetler, C. A., Di Girolamo, P., ... & Zoffoli, S. (2023). Satellite Lidar Measurements as a Critical New Global Ocean Climate Record. *Remote Sensing*, *15*(23), 5567.

Behrenfeld, M. J., Hu, Y., Bisson, K. M., Lu, X., & Westberry, T. K. (2022). Retrieval of ocean optical and plankton properties with the satellite Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) sensor: Background, data processing, and validation status. *Remote Sensing of Environment*, *281*, 113235.

Behrenfeld, M. J., Gaube, P., Della Penna, A., O'malley, R. T., Burt, W. J., Hu, Y., ... & Doney, S. C. (2019). Global satellite-observed daily vertical migrations of ocean animals. Nature, 576(7786), 257-261.

Behrenfeld, M. J., Hu, Y., O'Malley, R. T., Boss, E. S., Hostetler, C. A., Siegel, D. A., ... & Scarino, A. J. (2017). Annual boom–bust cycles of polar phytoplankton biomass revealed by space-based lidar. *Nature Geoscience*, *10*(2), 118-122.

Behrenfeld, M. J., Hu, Y., Hostetler, C. A., Dall'Olmo, G., Rodier, S. D., Hair, J. W., & Trepte, C. R. (2013). Space-based lidar measurements of global ocean carbon stocks. *Geophysical Research Letters*, *40*(16), 4355-4360.

Bisson, K. M., Boss, E., Werdell, P. J., Ibrahim, A., Frouin, R., & Behrenfeld, M. J. (2021a). Seasonal bias in global ocean color observations. Applied optics, 60(23), 6978-6988.

Bisson, K. M., Boss, E., Werdell, P. J., Ibrahim, A., & Behrenfeld, M. J. (2021b). Particulate backscattering in the global ocean: a comparison of independent assessments. *Geophysical research letters*, *48*(2), e2020GL090909.

Bisson, K. M., Werdell, P. J., Chase, A. P., Kramer, S. J., Cael, B. B., Boss, E., ... & Behrenfeld, M. J. (2023). Informing ocean color inversion products by seeding with ancillary observations. *Optics Express*, *31*(24), 40557-40572.

Bisson, K. M., & Cael, B. B. (2021). How are under ice phytoplankton related to sea ice in the Southern Ocean?. Geophysical Research Letters, 48(21), e2021GL095051.

Eidam, E., Bisson, K., Wang, C., Walker, C., & Gibbons, A. (2023). ICESat-2 and ocean particulates: Building a roadmap for calculating Kd from space-based lidar photon profiles.

Horvat, C., Bisson, K., Seabrook, S., Cristi, A., & Matthes, L. C. (2022). Evidence of phytoplankton blooms under Antarctic sea ice. *Frontiers in Marine Science*, *9*, 2154.

Hostetler, C. A., Behrenfeld, M. J., Hu, Y., Hair, J. W., & Schulien, J. A. (2018). Spaceborne lidar in the study of marine systems. *Annual review of marine science*, *10*, 121-147.

Huppenkothen, D., Arendt, A., Hogg, D. W., Ram, K., VanderPlas, J. T., & Rokem, A. (2018). Hack weeks as a model for data science education and collaboration. Proceedings of the National Academy of Sciences, 115(36), 8872-8877.

Lacour, L., Larouche, R., & Babin, M. (2020). In situ evaluation of spaceborne CALIOP lidar measurements of the upper-ocean particle backscattering coefficient. *Optics Express*, *28*(18), 26989-26999.

Lu, X., Hu, Y., Yang, Y., Bontempi, P., Omar, A., & Baize, R. (2020). Antarctic spring ice-edge blooms observed from space by ICESat-2. Remote Sensing of Environment, 245, 111827.

Lu, X., Hu, Y., Yang, Y., Neumann, T., Omar, A., Baize, R., ... & Winker, D. (2021). New Ocean Subsurface Optical Properties From Space Lidars: CALIOP/CALIPSO and ATLAS/ICESat-2. Earth and Space Science, 8(10), e2021EA001839.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., ... & Zwally, J. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation. Remote sensing of environment, 190, 260-273.

Parrish, C. E., Magruder, L. A., Neuenschwander, A. L., Forfinski-Sarkozi, N., Alonzo, M., & Jasinski, M. (2019). Validation of ICESat-2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance. Remote sensing, 11(14), 1634.

PyPI (2023). *Python Package Index*. Computer software. Python Software Foundation, Feb. 2023. <https://pypi.org/>

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., ... & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 1-10.

Scheick, J. et al., (2019). icepyx: Python tools for obtaining and working with ICESat-2 data. https://github.com/icesat2py/icepyx.

Scheick, J., Leong, W. J., Bisson, K., Arendt, A., Bhushan, S., Fair, Z., ... & Swinski, J. (2023). icepyx: querying, obtaining, analyzing, and manipulating ICESat-2 datasets. *Journal of Open Source Software*, *8*(84), 4912.

Salisbury, Joseph. "The Geosynchronous Littoral Imaging and Monitoring Radiometer (GLIMR): NASA's newest Ocean Color Mission." *44th COSPAR Scientific Assembly. Held 16-24 July*44 (2022): 95.

Tang, W., Llort, J., Weis, J., Perron, M. M., Basart, S., Li, Z., ... & Cassar, N. (2021). Widespread phytoplankton blooms triggered by 2019–2020 Australian wildfires. *Nature*, *597*(7876), 370-375.

Thomas, D., & Hunt, A. (2019). *The pragmatic programmer, 20th anniversary edition: Journey to mastery* (Second edition). Addison-Wesley.

Tucker, T., Giglio, D., Scanderbeg, M., & Shen, S. S. (2020). Argovis: a web application for fast delivery, visualization, and analysis of Argo data. *Journal of Atmospheric and Oceanic Technology*, *37*(3), 401-416.

United States White House Office of Science and Technology Policy (OSTP) and National Science and Technology Council (NSTC) (2023). FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research. https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/.

van Rossum, G., Warsaw, B., & Coghlan, N. (2001, July 5). *PEP 8 – Style Guide for Python Code*. Peps.Python.Org. https://peps.python.org/pep-0008/#introduction

Werdell, P. J., Behrenfeld, M. J., Bontempi, P. S., Boss, E., Cairns, B., Davis, G. T., ... & Remer, L. A. (2019). The Plankton, Aerosol, Cloud, ocean Ecosystem mission: status, science, advances. *Bulletin of the American Meteorological Society*, *100*(9), 1775-1794.