OOI 1.0 Data Evaluation Summary Report

By:
Leila Belabbassi
Lori Garzio

Affiliation:
Department of Marine and Coastal Sciences
Rutgers Center for Ocean Observing Leadership
School of Environmental and Biological Sciences
Rutgers The State University of New Jersey

Table of Contents

List of Figures

List of Tables

OOI 1.0 Data Evaluation Summary Report

Overview

The Ocean Observatories Initiative (OOI) is a program that uses science-driven platforms and sensor systems to measure oceanographic properties and processes from the seafloor to the air-sea interface. Since OOI becoming operational in 2016, data from two coastal arrays (CP, CE), one cabled array (RS), and four global arrays (GA, GI, GP, GS) have been made available through the OOI data portal (OOINET). This data delivery comes with a responsibility to inform users about data quality and availability, as well as known issues with instruments and datasets.

As part of the OOI Synthesis & Education project led by the Consortium for Ocean Leadership (COL), the Data Team at Rutgers University reviewed and evaluated the official datasets collected by the OOI from July 2013 to September 2018 (this time period is referred to as OOI 1.0). This effort supported two goals within the project, first provide feedback to the OOI Cyberinfrastructure system developers and the Marine Implementing Organizations (MIOs) regarding issues with data availability and quality, enabling the system administrators to fix data issues and/or communicate issues to users via annotations in the OOINET system. Secondly, identify and collate OOI datasets that could be used by educators in the classroom. This report only addresses tasks associated with the data evaluation process.

Instruments for Review

During OOI 1.0, the OOI contained 1,327 science and engineering instruments (as defined by individual reference designators). Of those, instruments were selected for this review if they contained science parameters, had data downloadable via OOINET, and had at least one deployment that was recovered before Sept. 30, 2018. Therefore, the total number of instruments flagged for review was 957 (Table 1). Many of the selected instruments had multiple deployments, some had multiple data streams, and uncabled instruments had multiple data delivery methods, i.e. via telemetry and through various recovered data methods. To limit the number of datasets to a manageable size, one data delivery method was selected for each instrument to review. For uncabled instruments, the preferred delivery method was recovered-instrument (when available) because this should be the most complete, high resolution dataset. If recovered-instrument was not available, recovered-host (from the Data Concentrator Logger) or telemetered data were reviewed.

Table 1. Number of Reviews by Array. CE: Coastal Endurance, CP: Coastal Pioneer, GA: Global Argentine Basin, GI: Global Irminger Sea, GP: Global Station Papa, RS: Cabled Array

Review Category	CE	CP	GA	GI	GP	GS	RS	Total
Science Instruments	247	186	128	128	85	116	67	**957**
Instrument-deployment-streams	1829	1184	302	426	253	242	216	**4452**

Data Evaluation Process

For each OOI 1.0 instrument flagged for review, the data evaluator:
- Downloaded data through OOINET via the Machine-to-Machine interface
- Analyzed the downloaded files using an automated set of tools to test the basic components of each dataset (Table 2)
- Imported the automated analysis summary into the Data Review Portal for each instrument
- Plotted the data and stored the plots in a publicly accessible website, organized by reference designator
- Completed the Human-In-The-Loop (HITL) review
  - Examined automated test output, system annotations, and data plots
  - Entered notes in each instrument report page in the Data Review Portal regarding issues with data quality
- Calculated final data ranges for each applicable dataset excluding data outside of global ranges and suspect data
- Each month, exported all review notes and sent to COL as a data report for further distribution to the OOI systems and operator teams

Schematic of Data Evaluation Process
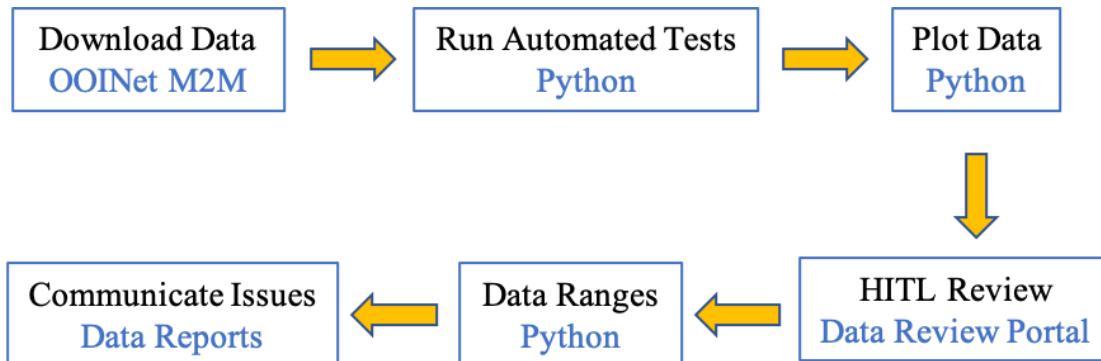
Table 2. Descriptions of automated data quality tests.

Test	Description
Deployment Days	Number of days the instrument was deployed.
File Days	Number of days for which there is at least 1 timestamp available for the instrument.
Start Gap	Number of missing days at the start of a deployment: comparison of the deployment start date to the data start date.
End Gap	Number of missing days at the end of a deployment: comparison of the deployment end date to the data end date.
Gap Count	Number of gaps within a data file (exclusive of missing data at the beginning and end of a deployment). Gap is defined as >1 day of missing data.
Gap Days	Number of days of missing data within a data file (exclusive of missing data at the beginning and end of a deployment).
Timestamps	Number of timestamps in a data file.
Sampling Rate	Sampling rates are calculated from the differences in timestamps. The most common sampling rate is that which occurs >50%.
Pressure Comparison	Instrument deployment depth defined in OOI's Asset Management system / average (for fixed instruments) or maximum (for mobile instruments) pressure calculated from data file after eliminating data outside of global ranges and outliers (3 standard deviations).
Time Order	Test that timestamps in the file are unique and in ascending order.
Valid Data	For each science variable, the binned percent of data that are not NaNs, fill values, outside global ranges, and outside 5 standard deviations. Bins: 99 = >99%, 95 = 95-99%, 75 = 75-95%, 50 = 50-75%, 25 = 25-50%, 0 = 0-25%. For example, {'99':4, '95':1} means 4 science variables have >99% valid data points, and 1 science variable has between 95-99% valid data points.
Missing Data	Test fails if data are available in another stream from a "non-preferred" delivery method, where the same data are not available in the preferred data stream. Summary provides the number of gaps and days of data that are missing in the preferred dataset that should be available.
Data Comparison	Compare data values with matching timestamps for science variables among all delivery methods.
Missing Coordinates	Check the coordinates in the data file against expected coordinates: obs, time, lat, lon, pressure (for instruments not located on a surface buoy)

Instrument Evaluation Summary Reports

Each reviewed instrument has a summary data report in the Rutgers University Data Review Database that includes:

1. General information about the instrument
2. Summary output from the automated data checks
3. Data downloaded and reviewed date
4. Link to descriptions of the automated data quality tests
5. Automated test notes
6. Links to data ranges (if applicable) and plots
7. Data Coverage table showing % data available for each science data stream for each OOI 1.0 deployment
8. Table of platform location differences among deployments

CTD (30 m)

OOI Site Page | Data Portal | Info | Report

(1)

Reference Designator	GP03FLMA-RIM01-02-CTDMOG040
Review Status	✔ Review Complete
Note	Deployment 2 recovered_inst data should be made available for download from uFrame. Deployment 5 data gap should be filled.

Depth	30m
Class	CTDMO (CTD Mooring (Inductive))
Make / Model	Sea-Bird / SBE 37IM

(3)
(4)

(2) **Dataset Reviews** *Last processed: 11/28/18, 5:20 PM*   ❶ QC Check Info

Dep.	Preferred Method	Stream	DD	FD	SG	EG	Gaps	GD	TS	Rate (s)	Pressure Comp.	Time Order	Valid Data	Missing Data	Data Comp.	Missing Coords.	Review
1	recovered_inst	ctdmo_ghqr_instru...	329	329	0	0	0	0	31,493	900	28 / 15	✔	✔	✔	✔	✔	Complete
2	telemetered	ctdmo_ghqr_sio_mu...	353	157	0	195	0	0	914	14,400	28 / 12	✔	✔	❶	❶	✔	Complete ★
3	recovered_inst	ctdmo_ghqr_instru...	389	389	0	0	0	0	37,234	900	48 / 33	✔	✔	✔	✔	✔	Complete
4	recovered_inst	ctdmo_ghqr_instru...	383	383	0	0	0	0	36,728	900	/ 42	✔	✔	✔	✔	✔	Complete
5	recovered_inst	ctdmo_ghqr_instru...	380	373	0	0	1	8	35,581	900	/ 40	✔	✔	❷	✔	✔	Complete ★

(5) **Test Notes**   **(6)** ❶ Data Ranges | Review Images

1. no other streams for comparison
2. fail: data found in another stream (gaps: [1] days: [9])

(7) **Data Coverage**

Deployment:	1	2	3	4	5
ctdmo_ghqr_instru...	100%		100%	100%	98%
ctdmo_ghqr_sio_mu...		44%			

(8) **Lat/Lon Differences (km)**

Deployment:	1	2	3	4	5
1	0.00				
2	0.62	0.00			
3	0.65	0.09	0.00		
4	1.16	0.56	0.51	0.00	
5	9.06	9.67	9.72	10.22	0.00

9. Composite data plot (if applicable) displaying data from one representative science parameter from the preferred data steam(s) for each deployment. Plot shows data cleaned of erroneous values after the HITL review
10. System annotations imported from the OOI Cyberinfrastructure
11. Review notes entered by the data reviewer during the HITL data review. These notes were exported monthly and shared with the MIOs in an effort to improve the quality of the OOI 1.0 datasets

Example Composite Data Plot ⓘ

(9)

GP03FLMA-RIM01-02-CTDMOG040
Deployments: [1, 2, 3, 4, 5]
2013-07-21T22:45:01 - 2018-07-27T22:00:01
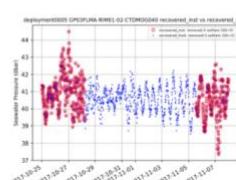
(10) **System Annotations** *Show*

Metadata	Start Date	End Date	Comment
🏷 GP03FLMA	7/21/13, 6:44 PM	6/14/14, 8:00 PM	Deployment 1: Mooring was deployed 10-15m shallower than planned. Telemetered data are not available for this deployment. Id: 150 By: lgarzio
🏷 GP03FLMA-RIM01-02-CTDMOG040	6/17/14, 2:00 AM	6/4/15, 8:00 PM	Deployment 2: Upon recovery, a shield that guards the sensor and holds the anti-fouling was missing. Id: 582 By: lgarzio

(11) **Review Notes**

Metadata	Start Date	End Date	Comment
☀ GP03FLMA	7/1/16, 1:07 AM	7/27/18, 10:14 PM	Deployments 4 and 5 don't have deployment depths defined in asset management. *By Lori Garzio, on 1/3/19*
☀ GP03FLMA-RIM01-02-CTDMOG040 Deployment: 2	6/17/14, 6:00 AM	6/5/15, 12:00 AM	recovered_inst data are not available for download from uFrame. There is an entry in the ingest csv indicating that the raw data file is available. This should be investigated and the recovered_inst data should be made available for download (or the system should have an annotation indicating why the recovered_inst data are not available). *By Lori Garzio, on 1/4/19*
☀ GP03FLMA Deployment: 5	7/13/17, 11:00 PM	7/27/18, 10:14 PM	The mooring was deployed 9 - 10 km from the location of the previous 4 deployments of this mooring. If this location is correct, the platform should be annotated to inform the users as to why the deployment location changed. *By Lori Garzio, on 1/4/19*
☀ GP03FLMA-RIM01-02-CTDMOG040 Deployment: 5	10/28/17, 5:15 AM	11/5/17, 1:00 PM	There is a data gap in the recovered_inst data stream from 2017-10-28T05:15:01 to 2017-11-05T13:00:01, however these data are available in the recovered_host data stream. The recovered_inst data should be filled in. See Redmine ticket 13752. *By Lori Garzio, on 1/4/19*

Summary of Data Issues

This summary was collated using Helpdesk tickets submitted during this project via Redmine, as well as the notes recorded in the Rutgers University Data Review Database during the HITL data review. The notes were exported and shared with the Marine Implementing Organizations monthly in an effort to improve the quality of the OOI 1.0 datasets, and these reports are stored in the data review section of the OOI Data Labs Github repository.

1. General Issues

1.1 Informing end-users of data quality, operational, and cyberinfrastructure issues

Informing end-users of data quality and operational issues is a critical responsibility of OOI data delivery, however annotations in the OOI system are often lacking or absent. During the data review it was observed on many occasions that large gaps in data availability were not annotated or explained. Table 3 includes a short example list of large gaps that were not annotated, a full list can be found in the Review Notes section of the Data Review Portal, or in the monthly data review reports. Some of the gaps in data availability are presumably known by the operators, and any information regarding gaps should be communicated to the public.

Table 3. Short example list of large data gaps that are not annotated.

Reference Designator	Issue/Gap not annotated
GA01SUMO-RII11-02-FLORDG031	193 missing days at the end of deployment 2
CE09OSSM-MFD35-06-PHSEND000	159 missing days at the end of deployment 1
CP04OSSM-MFD37-03-CTDBPE000	42 missing days at the end of deployment 4
CE06ISSM-SBD17-06-CTDBPC000	13 – 137 missing days at the end of deployments 3-7
RS03ASHS-MJ03B-07-TMPSFA301	Changes in temperature recorded by several thermistors after an unexplained data gap

1.2 Annotation review and updates

Many annotations within the system should be reviewed and updated. For example, there are several duplicated annotations that should be deleted (e.g. RS01SBPS-SF01A-4A-NUTNRA101 annotation ID 1370 duplicates 572). There are also annotations from several years ago that require resolution – e.g., an investigation is pending (CE04OSPS-SF01B-2A-CTDPFA107 annotation ID 1045-1046 and 1064), a correction needs to be applied (e.g. CE02SHBP-LJ01D-06-CTDBPN106 annotation ID 201), or an algorithm is currently under review and all data are considered incorrect (CP02PMCI-WFP01-01-VEL3DK000 annotation ID 1402). If these annotations are outdated, they need to be updated so users have the most current information regarding these datasets. Lastly, deviations from planned sampling rates should be annotated (e.g. CP01CNSM-RID27-03-CTDBPC000 deployments 5-7). All annotations that were identified as requiring a review and possible update can be found in the [Review Notes](#) section of the Data Review Portal, or in the monthly [data review reports](#). Moving forward, it would be helpful for operators to have the ability to flag annotations that need to be reviewed and modified.

1.3 Missing data

Numerous instrument deployments for each array are not available for download from the OOI Cyberinfrastructure (Table 4) and there are no annotations to explain why. These data either need to be ingested or a reason for their absence provided (full list: **Tables S1.3a - S1.3d**).

Table 4. Number of instruments missing for at least one deployment/delivery method, by array

Array	# Instruments	Details
Endurance	78	Table S1.3a
Pioneer	42	Table S1.3b
Global	49	Table S1.3c
Cabled	15	Table S1.3d

1.4 NetCDF file coordinates missing or incorrectly assigned

An instrument's location in three-dimensional space (latitude, longitude and depth) is critical for data use/interpretation and should be provided as coordinates in every NetCDF file generated by the OOI system. Table 5 highlights instrument types with NetCDF files that are currently missing one or all of these coordinates (note: the pressure coordinate is not expected on instruments connected to a surface buoy, all instruments listed below refer to those deployed sub-surface):

Table 5. Missing coordinates in NetCDF files by instrument type

Instrument Type	Coordinate missing
Cabled ADCP & VADCP	Pressure, latitude, longitude
Cabled FLORD	Pressure, latitude, longitude
D1000	Pressure
HPIES	Pressure, latitude, longitude
NUTNR	Pressure
OPTAA	Pressure
PCO2W	Pressure
PHSEN	Pressure
SPKIR	Pressure
THSPH	Pressure
TMPSF	Pressure
TRHPH	Pressure

The pressure coordinate is a critical piece of information that is required for all instruments deployed sub-surface. This coordinate is not included in NetCDF files in some instances where the collocated CTD data are not available (e.g. CP01CNSM-RID27-04-DOSTAD000 deployments 2, 3, 5, and 6). This coordinate should always be included in the files, and if the pressure from the CTD is not available, the array should be filled with NaNs or fill values. A full list of instances where the pressure coordinate is missing from NetCDF files can be found in the Review Notes section of the Data Review Portal, or in the monthly data review reports.

To exacerbate the problem, the system currently only pulls collocated data for the same delivery method when calculating L2 data products. For example, for most DOSTAs and FLORTs, the delivery method is recovered_host. If the collocated recovered_host CTD data are not available, the L2 data products are not calculated. The system should be flexible enough to use a different delivery method (e.g. recovered_inst) for L2 data product calculations if data from the preferred delivery method is not available.

1.5 Data only available in non-preferred data stream

The most complete, high-resolution "preferred" dataset for all uncabled instruments should be the recovered_inst dataset which is downloaded directly from the instrument upon recovery. When this delivery method is not available (e.g., the instrument does not record internally or the instrument was lost or damaged), the recovered_host delivery method would be the "preferred" data stream. However, in multiple cases, portions of deployment data are available in the telemetered and/or recovered_host stream, and the same data are not available in the preferred recovered_inst data stream (Fig. 1, full list: **Tables S1.5a - S1.5c**).
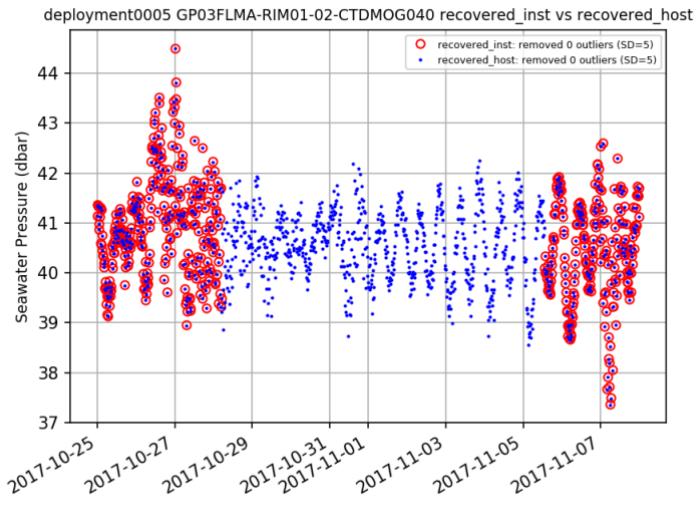
Figure 1. Example where data are available in a "non-preferred" data stream (recovered_host: blue dots), and not available in the "preferred" data stream (recovered_inst: red circles) for part of the deployment.

1.6 Incomplete and missing shipboard data
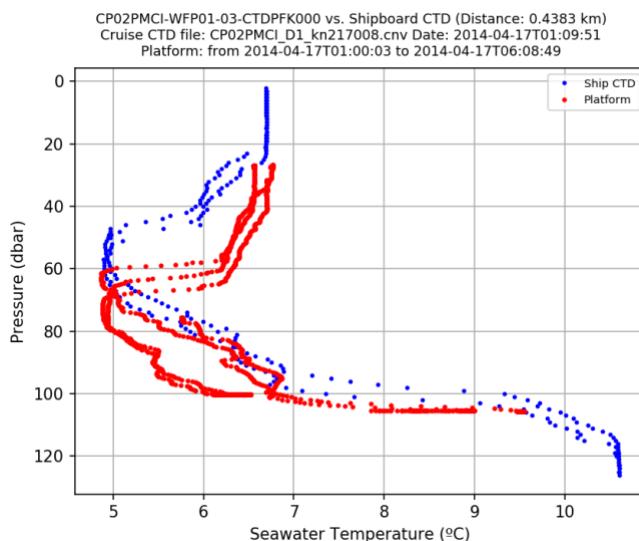
Figure 2. Example comparison of CTD data from deployment 1 of the Pioneer Central Inshore Profiler Mooring with a shipboard CTD cast done at the time of deployment.

Supplemental data collected on OOI maintenance cruises (e.g., shipboard CTD casts, water sampling data) are crucial for ground-truthing measurements from OOI platforms, as well as post-processing to account for sensor drift. Comparisons of data collected from OOI platforms with shipboard CTD casts were attempted during this review, however the majority of Endurance and Cabled shipboard data were unavailable at the time of analysis (March 2019) and some of the Pioneer and Global data were also missing. From the comparisons that were conducted (mostly Pioneer and Global data), differences between the platforms and the shipboard CTD casts were often substantial (Fig. 2). Further analyses using the shipboard bottle data should be conducted. This requires all of the shipboard data be made available to the public as soon as possible after maintenance cruises are completed and missing historical data from past maintenance cruises be uploaded.

1.7 Missing and incorrect global ranges

Each instrument and science parameter should have global ranges that are used in the OOI Global Range Test as part of the defined QA/QC process. Several instruments and parameters were identified that either had missing global ranges (Table 6), or the global ranges looked incorrect and needed to be reviewed (Table 7).

Table 6. Instruments and parameters with missing global ranges

Instrument Type	Parameter(s)
D1000	All science parameters
FDCHP	fdchp_a_fluxhot, fdchp_a_fluxmom_alongwind, and fdchp_a_fluxmom_crosswind
FLORTO	sci_bb3slo_b470_units, flort_o_bback_b470_total, flort_o_bback_b532_total, sci_bb3slo_b660_units, and flort_o_bback_b660_total
HPIES	hpies_pressure_L1
METBK	met_windavg_mag_corr_east, met_windavg_mag_corr_north, met_heatflx_minute, met_latnflx_minute, met_netlirr_minute, met_sensflx_minute, met_buoyfls, met_buoyflx, met_frshflx, met_heatflx, met_latnflx, met_mommflx, met_rainflx, met_sensflx, met_netsirr_hourly
PRESF	scaled_wave_burst_seafloor_pressure
TRHPH	resistivity_temp_degc, vent_fluid_orp and trhph_thermistor_temp
WAVSS	wavss_a_dcl_statistics mean_spread, wavss_a_dcl_mean_directional_recovered wavss_a_directional_frequency, and wavss_a_dcl_non_directional_recovered wavss_a_non_directional_frequency
CE02SHSP-SP002	All instruments and science parameters
CP01CNPM	All instruments and science parameters
CP05MOAS-PG564	All instruments and science parameters
CP05MOAS-PG583	All instruments and science parameters
Global DOSTAs on RII11	dosta_analog_tc_oxygen
Global CTDBPs on RII11	ctdbp_cdef_instrument_recovered ctdbp_seawater_temperature, ctdbp_seawater_pressure, and ctdbp_seawater_conductivity

Table 7. Global ranges that need to be reviewed and revised

Instrument Type	Parameter(s)
ADCPT	temperature
HPIES	hpies_bliley_frequency
METBK	met_current_speed, met_relwind_speed, met_wind10m, shortwave_irradiance, and met_netsirr
PARADM	sci_bsipar_par
Cabled Deep profilers	conductivity_millisiemens
Global Irminger VELPTB	pressure_mbar

1.8 Incomplete local ranges

The Local Range Test, or an equivalent version such as a Climatology Test, is often required or strongly recommended for real-time quality control of oceanographic data to conform to IOOS QARTOD standards. The Local Range Test is one of the core Automated QC Algorithms listed in the quality control section of the OOI website, however local ranges for the majority of OOI instruments are incomplete. For the early OOI deployments, a sufficient time-series of data to establish the local ranges did not exist. Now that the many OOI instruments have been deployed at the same location for several years, local ranges can be computed from the collected datasets and used in the Local Range Tests. After excluding data that were suspect, outside of global ranges, and outside of a specified standard deviation, data ranges were calculated for each science variable in the applicable datasets for the OOI 1.0 timeframe as part of this review. These data ranges can be used to inform the local ranges.

2. Instrument-specific issues

2.1 FDCHP

Multiple FDCHP parameters are currently not being calculated (see Redmine 13402 for details, submitted 6/1/2018). When this issue is resolved, the annotations on the FDCHP datasets need to be updated.

2.2 FLORT

The pressure coordinate for every FLORT on Wire-Following Profilers is *pressure_depth*, but it always contains an array of fill values rather than valid pressure data. Another pressure variable is provided in the files called *int_ctd_pressure* that contains valid pressure data – this variable should either be the pressure coordinate, or the data should be populated in the *pressure_depth* coordinate. In addition, the variable *pressure_depth* is an array of fill values for every deployment of all stationary FLORTs.

The variable *total_volume_scattering_coefficient* (and *optical_backscatter*) on several instruments shows a repetitive pattern of increasing values to the upper detection limit of the instrument and then dropping to zero. Annotation IDs 97 and 98 on CE06ISSM-RID16-02-FLORTD000 describe the issue and indicate that the "investigation is ongoing". The same pattern is observed on several other instrument deployments. If this is a known issue these datasets should be annotated to inform users (Table 8, Fig. 3).

Table 8. Instrument deployments showing a repetitive pattern of *total_volume_scattering_coefficient* increasing to the upper detection limit and then dropping to zero.

Reference Designator	Deployment
CE06ISSM-RID16-02-FLORTD000	1
CE01ISSM-RID16-02-FLORTD000	2, 7
CE02SHSM-RID27-02-FLORTD000	6
CE04OSSM-RID27-02-FLORTD000	4
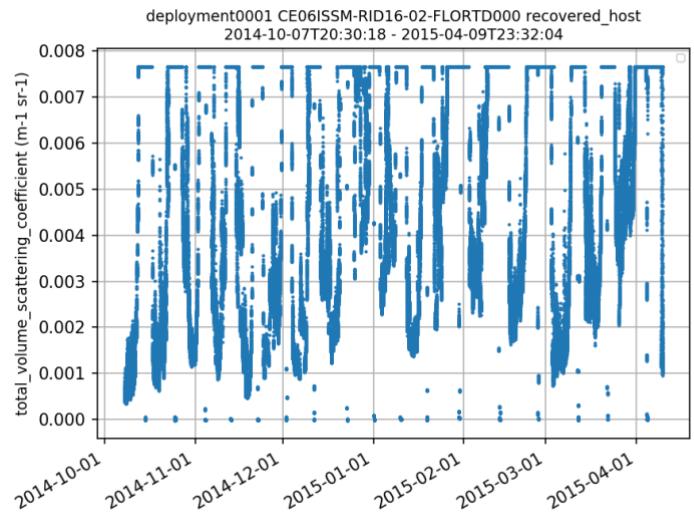
Figure 3. Timeseries plot of CE06ISSM-RID16-02-FLORTD000 showing a repetitive pattern of total_volume_scattering_coefficient increasing to the upper detection limit and then dropping to zero.

2.3 METBK

For deployment 2 of GS01SUMO-SBD11-06-METBKA000, *shortwave_irradiance* and *met_netsirr* minimum nighttime values jumped from ~0 W m$_{-2}$ to ~13-14 W m$_{-2}$ and remained there through the rest of the deployment (Redmine ticket # 12543). The issue was investigated and a correction was provided for this dataset (see annotation ID 1522). Several other instrument deployments appear to have a similar issue (Table 9, Fig. 4), all of which should undergo the same investigation and dataset corrections should be provided to users.

Table 9. Unreasonable jump in minimum nighttime *shortwave_irradiance* and *met_netsirr* values for several METBK instrument deployments.

Reference Designator	Deployment
GI01SUMO-SBD11-06-METBKA000	3
CE02SHSM-SBD11-06-METBKA000	1
CE07SHSM-SBD11-06-METBKA000	1
CE09OSSM-SBD11-06-METBKA000	1
CP01CNSM-SBD12-06-METBKA000	1
CP03ISSM-SBD11-06-METBKA000	3
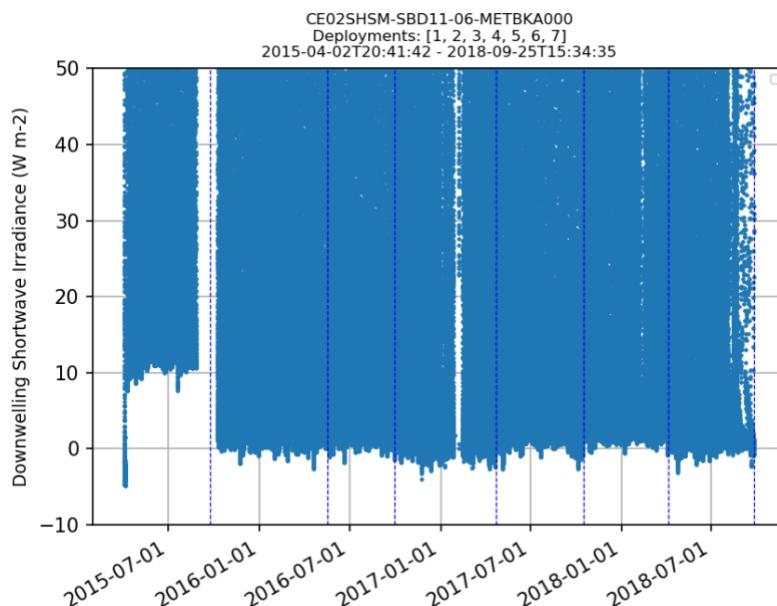
Figure 4. Example showing the jump in minimum values for Downwelling Shortwave Irradiance (*shortwave_irradiance*) at the beginning of deployment 1 for CE02SHSM-SBD11-06-METBKA000.

For several METBK deployments where no CT data are expected (e.g. the CT sensor was not connected correctly), *sea_surface_temperature* values were filled in as "-5", *sea_surface_conductivity* values as "0.0", and *met_salsurf* values as "0.0" (Table 10, Fig. 5). The defined fill values for all of these parameters is "-9999999". In addition, other variables appear to be using these data in their calculations, which is creating data products that should fail QC. For example, Sensible Heat Flux (*met_sensflx_minute*) uses *sea_surface_temperature* (PD1056) in its calculation. Sensible Heat Flux is substantially different for each deployment listed in Table 10 (e.g., Fig. 6), which suggests that the algorithm is erroneously using a value of -5.0 for *sea_surface_temperature* to calculate Sensible Heat Flux for these deployments.

Table 10. METBK instrument deployments where *sea_surface_temperature* = -5, *sea_surface_conductivity* = 0.0, *met_salsurf* = 0.0, and these values are being used in calculations of other data products.

| Reference Designator	Deployment
CE02SHSM-SBD11-06-METBKA000	5
CE09OSSM-SBD11-06-METBKA000	2
CP04OSSM-SBD11-06-METBKA000	1
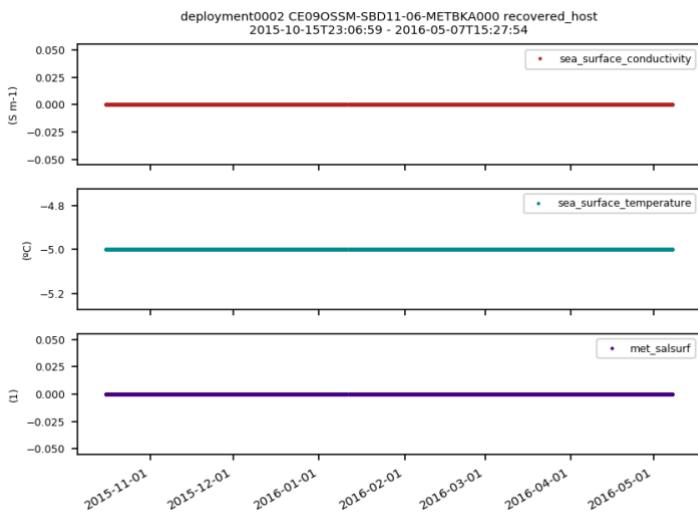
Figure 5. Deployment 2 of CE09OSSM-SBD11-06-METBKA000 where sea_surface_temperature = -5, sea_surface_conductivity = 0.0 and met_salsurf = 0.0.
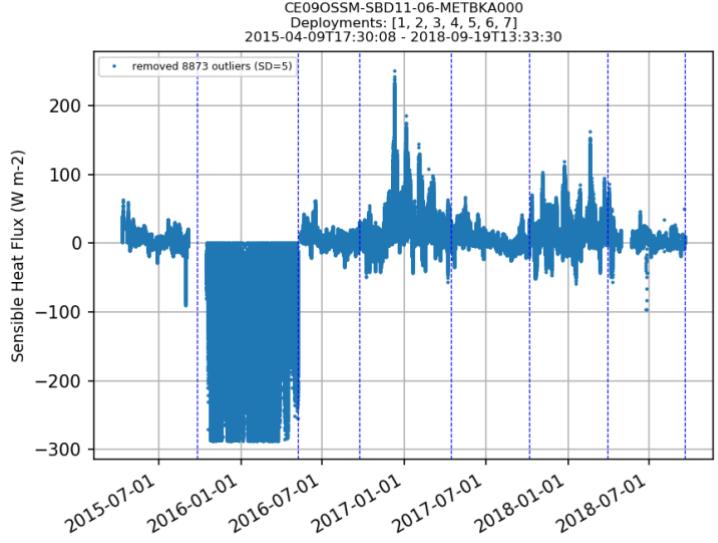

Figure 6. Sensible Heat Flux (met_sensflx_minute) for deployment 2 of CE09OSSM-SBD11-06-METBKA000 is substantially different compared to the other deployments of this instrument.

14

2.4 NUTNR

The annotation on all NUTNR instruments that were switched to the SUNA model in 2017 currently reads: "The Satlantic ISUS instrument has been discontinued, and all OOI ISUS units have been converted to the Sea-Bird SUNA model. A new data parser is in development, and any resulting data gaps will be filled once the parser has been delivered and the data are processed." This annotation needs to be updated when the new parser is implemented. In addition, all old NUTNR data that came from the problematic ISUS model should be annotated regarding the known issues with that model.

2.5 OPTAA

A substantial percentage of the *optical_absorption* and *beam_attenuation* data are unreasonable (Fig. 7). According to Roesler and Barnard[1], "absorption meters are highly prone to biofouling, particularly biofilms which not only attenuate the collimated beam but also impact the scattering properties of the optical surfaces and tubes. In productive coastal waters biofouling can have significant impacts (i.e. 10% of the signal) within one to two weeks". As such, these datasets should be annotated and further investigation of the impacts of biofouling on data quality needs to be conducted.

The wavelength variables (*wavelength_a* and *wavelength_c*) are arrays of fill values for some deployments (Table 11). This needs to be investigated and the wavelength variables need to be filled with valid data as they are crucial for data analysis and interpretation.

Table 11. OPTAA instrument deployments where *wavelength_a* and *wavelength_c* are arrays of fill values.

Reference Designator	Deployment
CE02SHBP-LJ01D-08-OPTAAD106	1
CE06ISSP-SP001-04-OPTAAJ000	3
CE07SHSM-RID27-01-OPTAAD000	4
CP01CNSM-RID27-01-OPTAAD000	7, 8
CP01CNSM-MFD37-01-OPTAAD000	6, 7
CP03ISSM-MFD37-01-OPTAAD000	6
CP04OSSM-RID27-01-OPTAAD000	5
CP04OSSM-MFD37-01-OPTAAD000	5
GA01SUMO-RID16-01-OPTAAD000	2
GI01SUMO-SBD12-01-OPTAAD000	4
RS01SBPS-SF01A-3B-OPTAAD101	2

[1] Roesler CS, Barnard AH. 2013. Optical proxy for phytoplankton biomass in the absence of photophysiology: Rethinking the absorption line height. Methods in Oceanography, 7, p 79-94.

Data Issues – Instrument-specific
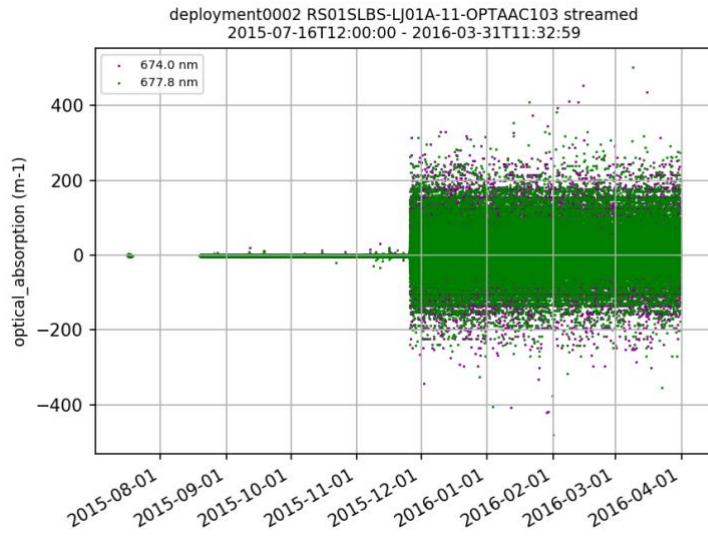
Figure 7. Optical Absorption values degrade over time for RS01SLBS-LJ01A-11-OPTAAC103.

2.6 PCO2A

For several deployment transitions, values at the end of one deployment are substantially different than the values at the beginning of the next deployment (Fig. 8), suggesting some sort of sensor drift or issue over time. If there is a known issue with these instruments, the datasets should be annotated.
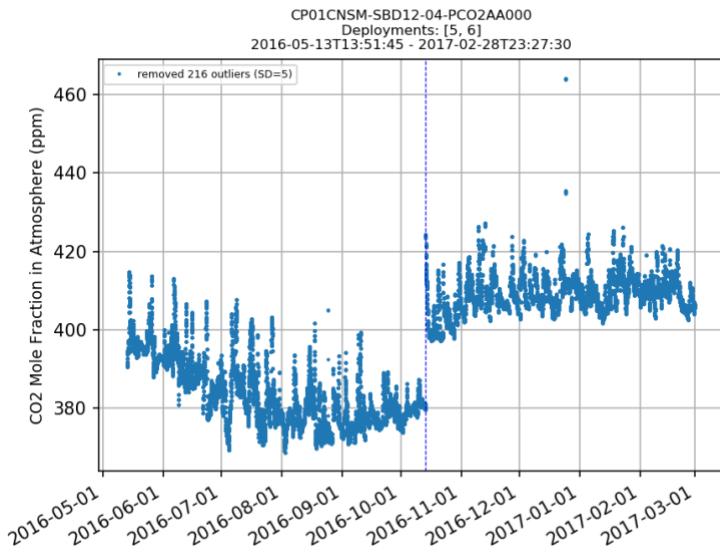
Figure 8. Values for CO2 Mole Fraction in Atmosphere from CP01CNSM-SBD12-04-PCO2AA000 are approximately 20 ppm lower at the end of deployment 5 compared to the values at the beginning of deployment 6.

Recovered data were not available for download for four deployments of Pioneer and Endurance PCO2A data. The issue was discovered while attempting to create Jupyter Notebook tutorials for educators using PCO2A data. A Helpdesk ticket was submitted regarding the missing data (#14529). In November 2019, the Marine Implementing Organizations (MIOs) ingested these data into the OOI cyberinfrastructure or annotated the datasets to explain why the data are not available for download and the issue was resolved.

CO$_2$ Flux (*pco2_co2flux*) currently is not being provided by the system for Pioneer and Global datasets (see Redmine ticket #14527 for details). In order to create pCO$_2$ air-sea flux tutorials for educators, this variable was calculated externally using the pco2_co2flux function in co2_functions from the OOI ion_functions. When comparing these results to those provided by the OOI system for Endurance datasets, it was observed that the OOI cyberinfrastructure was inverting the flux direction (Fig. 9). A Helpdesk ticket was submitted (#14531). The MIOs determined that the inputs to the function were reversed for *pco2a* and *pco2w*, and the issue was fixed in December 2019.
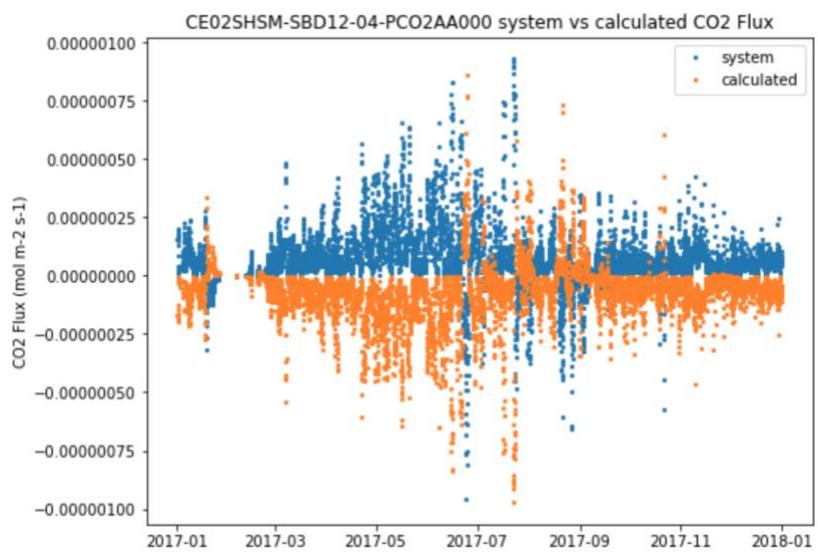
Figure 9. CO2 Flux (pco2_co2flux) provided by the OOI cyberinfrastructure (blue) compared to CO2 Flux calculated externally (orange) before the resolution in December 2019.

2.7 PCO2W

Recovered_inst data are not available for download for any uncabled instrument. According to the ingest .csv files, there are raw recovered_inst data files available – these data should be ingested.

At the time of the review, the majority (>50%) of *pco2_seawater* values for every deployment of every instrument were outside of global ranges, which were originally 100-600 µatm. The global ranges were updated in November 2019 to 200-2000 µatm, which will resolve the high percentage of data outside of global ranges.

2.8 PRESF

The second dimension for the *wave_burst* seafloor pressure should be described in the documentation or global attributes. It is currently an index from 0-1023 and is unclear to what this is referring.

Data Issues – Instrument-specific

2.9 SPKIR

For several deployments of every SPKIR, a substantial percentage (10% - 75%) of data are slightly negative and fail the Global Range QC test (Fig. 10).

2.10 VEL3D

The variable *vel3d_k_pressure* is an array of fill values for most deployments of uncabled instruments. This variable should contain valid pressure data. A full list of the instrument deployments affected can be found in the [Review Notes](#) section of the Data Review Portal, or in the monthly [data review reports](#).
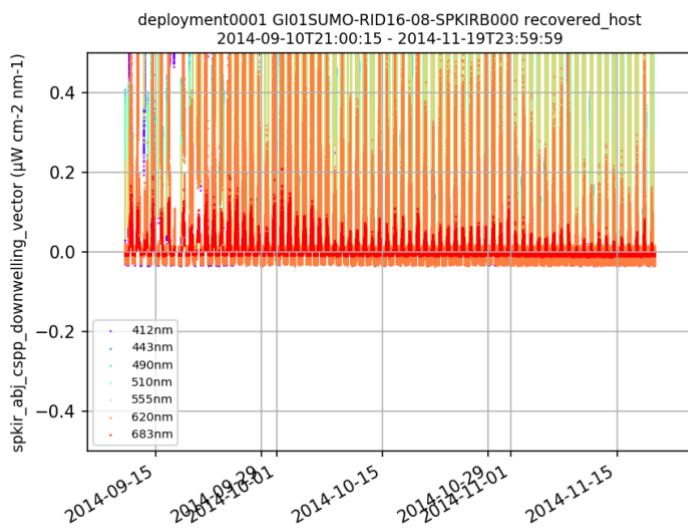
Figure 10. For deployment 1 of GI01SUMO-RID16-08-SPKIRB000, 40% of the values for every Downwelling Spectral Irradiance wavelength are slightly negative and fail the Global Range QC test.

2.11 WAVSS

Global and Pioneer WAVSS instruments were incorrectly configured for several deployments (see Redmine ticket #13404) and were annotated to alert users. This issue was first noticed by an end-user because the Number of Zero Crossings for one deployment were substantially different compared to other deployments at the same location. Patterns in the data suggest there are three deployments of Endurance instruments that have the same issue (Fig. 11, Table 12). This should be investigated, and the datasets should be annotated if there was a configuration mistake.

Table 12. Endurance WAVSS instrument deployments where the Number of Zero Crossings are substantially lower compared to all of the other deployments, suggesting a configuration error.

Reference Designator	Deployment
CE02SHSM-SBD12-05-WAVSSA000	4
CE07SHSM-SBD12-05-WAVSSA000	1
CE09OSSM-SBD12-05-WAVSSA000	5
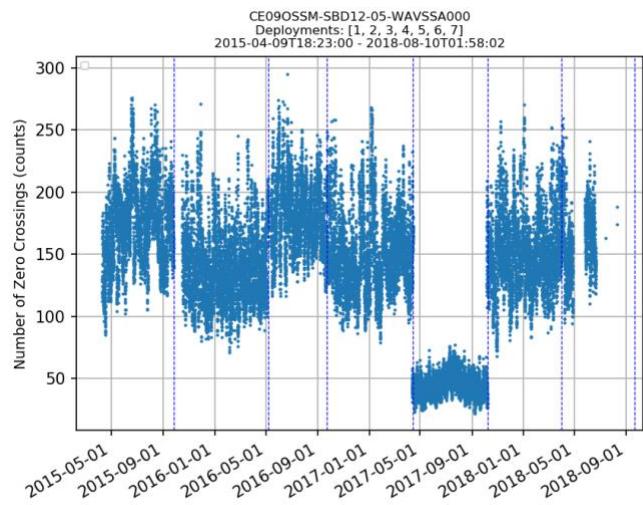
Figure 11. The Number of Zero Crossings for deployment 5 of CE09OSSM-SBD12-05-WAVSSA000 are substantially lower compared to all other deployments, suggesting a configuration error.

3. Endurance Array Issues

3.1 ADCP

The pressure variable for ADCPs is an array of fill values – this variable should contain valid pressure data.

3.2 DOSTA

Daily spikes in Dissolved Oxygen several months into each deployment suggest biofouling is an issue for these instruments (Fig. 12). These suspect data should be annotated.

For several instrument turnovers, Dissolved Oxygen values at the end of one deployment do not resemble the values at the beginning of the next deployment (Fig. 13), which could be indicative of sensor drift or another issue. Further analysis using shipboard data is required, and datasets should be annotated if there is a known issue. A full list of the instrument deployments affected can be found in the Review Notes section of the Data Review Portal, or in the monthly data review reports.
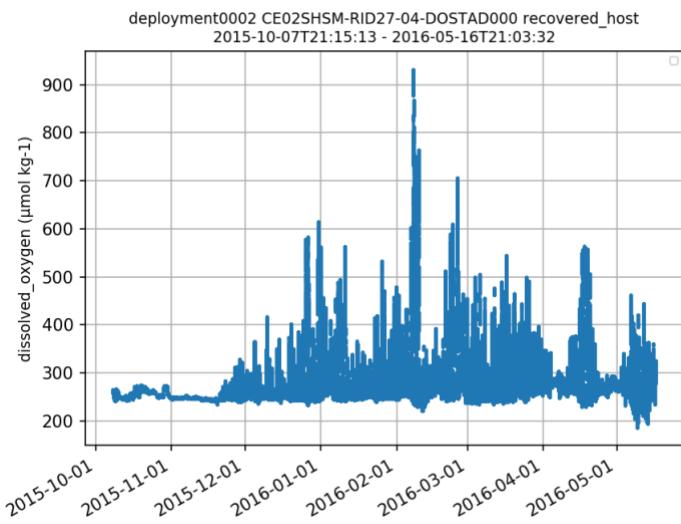
Figure 12. Example of unreasonable Dissolved Oxygen data, likely due to biofouling.
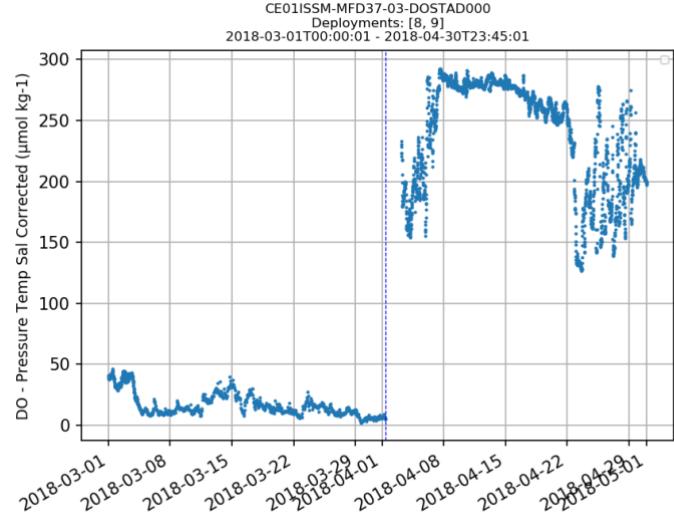
Figure 13. Dissolved Oxygen values at the end of deployment 8 are <50 µmol kg-1 and jump to 150-200 µmol kg-1 at the beginning of deployment 9 for CE01ISSM-MFD37-03-DOSTAD000.

3.3 METBK

Starting in the summer of 2018, all four Endurance METBKs exhibited varying degrees of suspect data for *sea_surface_conductivity*, *sea_surface_temperature*, and variables related to longwave irradiance (Fig. 14). If there is a known issue with these instruments at this time, the datasets should be annotated. In addition, other general data issues should be annotated (e.g., wide ranges in Sea Surface Conductivity) if these are known issues (Fig. 15). Finally, for some deployments, *eastward_velocity* and *northward_velocity* do not have any associated metadata, so it is unclear that these data come from the collocated VELPT (Fig. 16). A full list of the instrument deployments affected can be found in the Review Notes section of the Data Review Portal, or in the monthly data review reports.
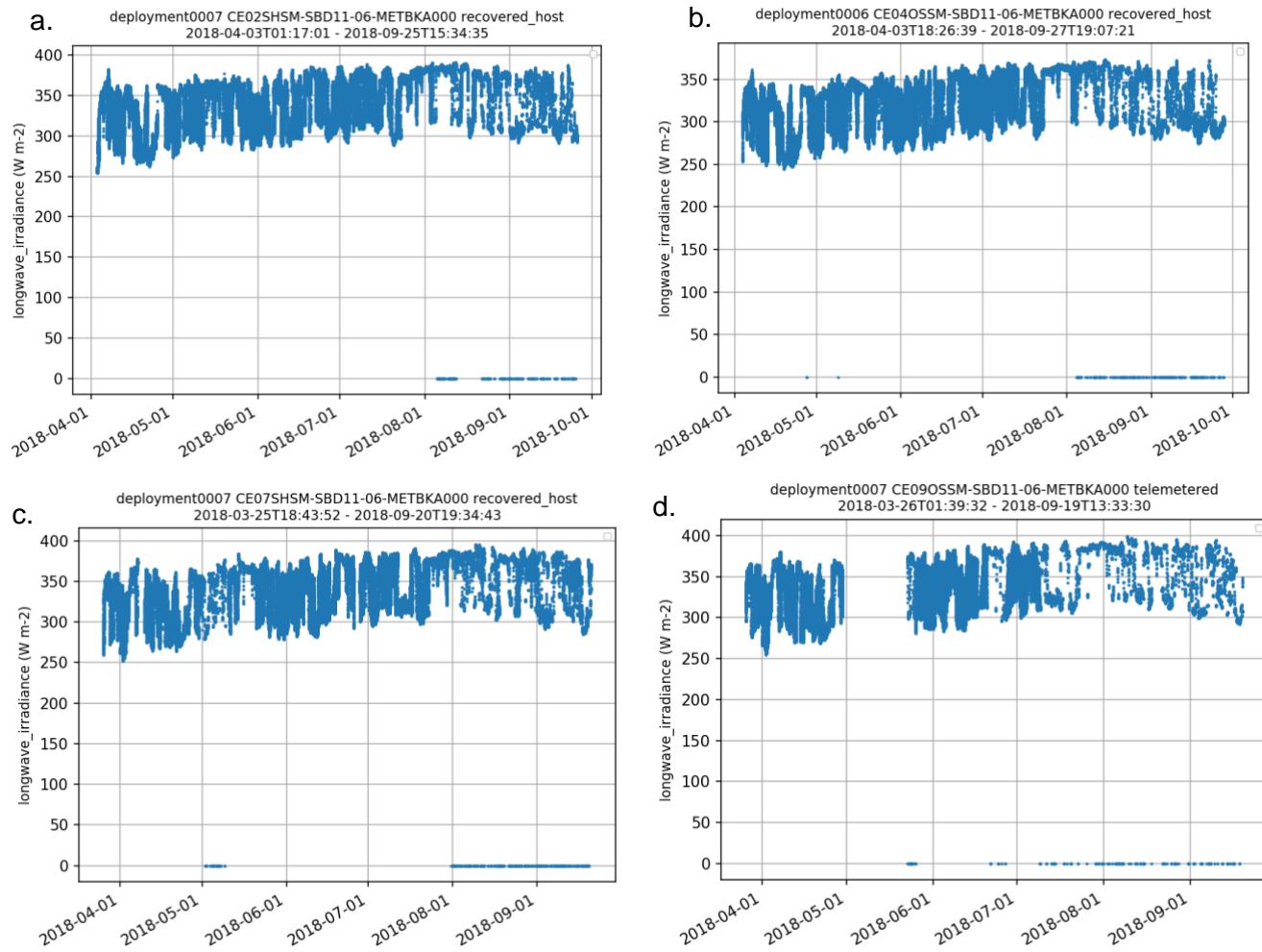
Figure 14. Suspect Endurance METBK data in the summer/fall of 2018 for (a) CE02SHSM, (b) CE04OSSM, (c) CE07SHSM, and (d) CE09OSSM.
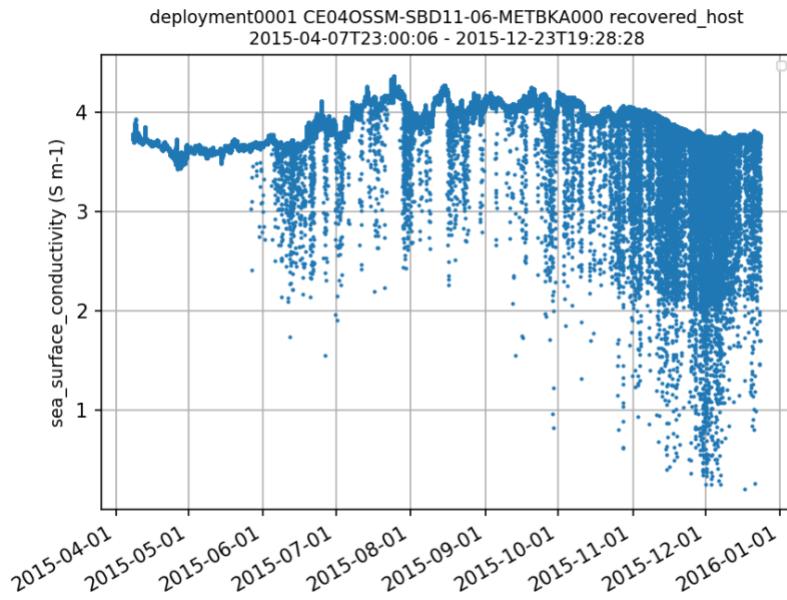
Figure 15. Issue with CE04OSSM-SBD11-06-METBKA000 deployment 1 Sea Surface Conductivity that requires an annotation.
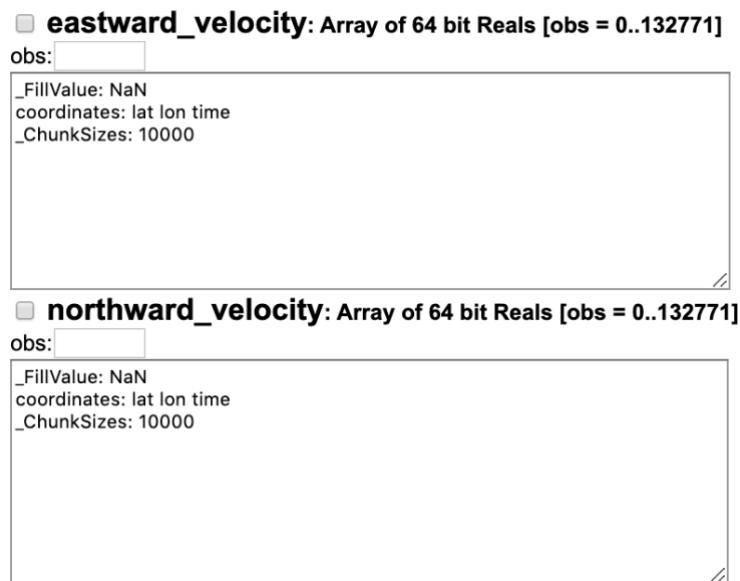
Figure 16. Screen shot of eastward_velocity and northward_velocity from the OPENDAP link for the deployment 4 CE07SHSM-SBD11-06-METBKA000 recovered_host metbk_a_dcl_instrument_recovered showing the lack of any identifying metadata.

3.4 Surface Piercing Profilers

For OPTAAs on Surface Piercing Profilers, the coordinate *pressure_depth* is an array of zeros. *int_ctd_pressure* contains valid pressure data and should be used to populate the coordinate variable.

For most deployments of NUTNRs on Surface Piercing Profilers, *ctd_temp* = -1, *ctd_psu* = 0 or -1, *ctd_dbar* = -1 (Fig. 17), and for some deployments >50% of *nitrate_concentration* and *nutnr_nitrogen_in_nitrate* values are negative (Fig. 18). This issue should be investigated and the datasets should be annotated if there are known issues.
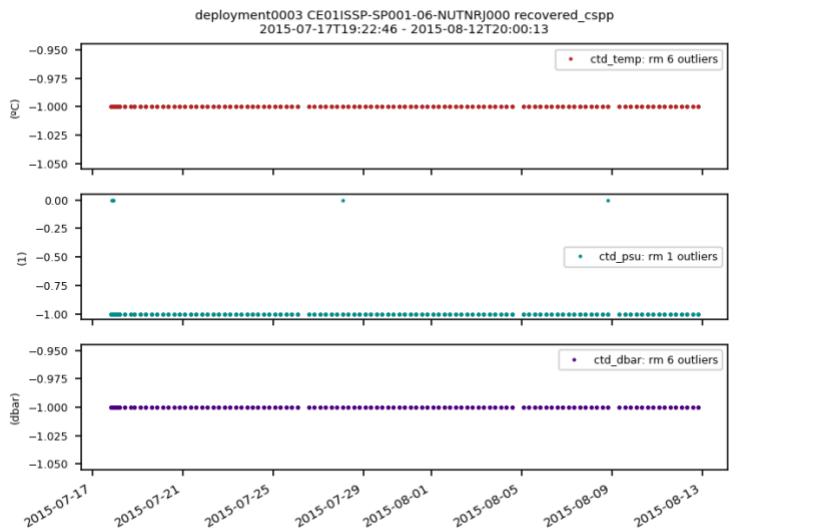
Figure 17. Incorrect values for ctd_temp, ctd_psu, and ctd_dbar for deployment 3 of CE01ISSP-SP001-06-NUTNRJ000
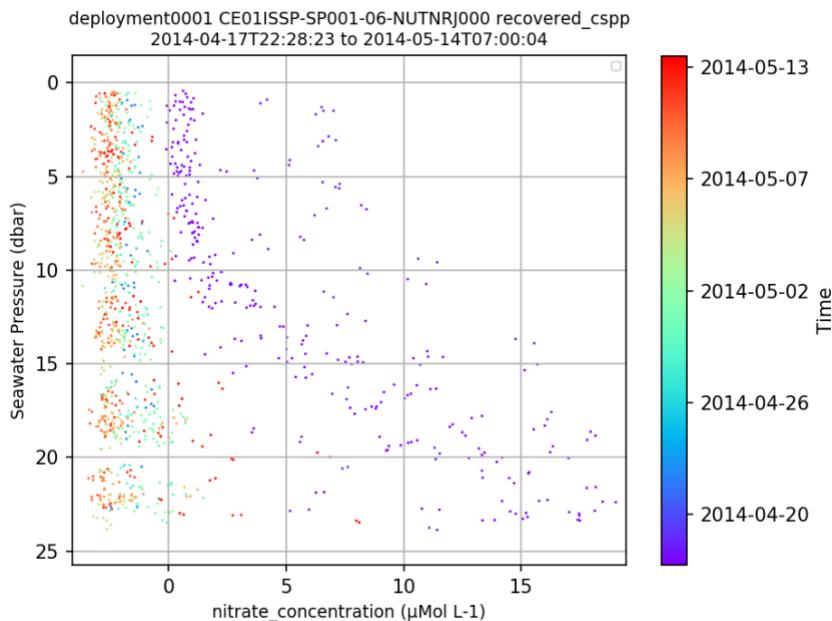
Figure 18. 70% of nitrate_concentration values are negative for deployment 1 of CE01ISSP-SP001-06-NUTNRJ000.

4. Pioneer and Global Array Issues

4.1 ADCP

Telemetered and recovered_host ADCP NetCDF files are missing two key variables needed for data interpretation: *percent_bad_beams* and *bin_depths* (Table 13).

4.2 PCO2A

The variable *pco2_co2flux* is an array of fill values for every deployment of the Pioneer and Global PCO2A instruments (Table 14). According to the provenance.json files, there is an error in the pco2_co2flux function in ion_functions.data.co2_functions. A Helpdesk ticket (#14527) was submitted in November 2019 and the issue is still under investigation as of April 2020.

Table 13. Pioneer and Global instruments where the telemetered and recovered_host NetCDF files are missing percent_bad_beams and bin_depths.

Reference Designator
CP01CNSM-MFD35-01-ADCPTF000
CP02PMCI-RII01-02-ADCPTG010
CP02PMCO-RII01-02-ADCPTG010
CP02PMUI-RII01-02-ADCPTG010
CP02PMUO-RII01-02-ADCPSL010
CP03ISPM-RII01-02-ADCPTG010
GA01SUMO-RII11-02-ADCPSN010

Table 14. Pioneer and Global PCO2A instruments that do not have valid pco2_co2flux data.

Reference Designator
CP01CNSM-SBD12-04-PCO2AA000
CP03ISSM-SBD12-04-PCO2AA000
CP04OSSM-SBD12-04-PCO2AA000
GA01SUMO-SBD12-04-PCO2AA000
GI01SUMO-SBD12-04-PCO2AA000
GS01SUMO-SBD12-04-PCO2AA000

4.3 METBK calibration coefficient

The calibration coefficient CC_use_velpt was changed from 1 to 0 in July 2019 as part of a metadata update for all Pioneer and Global METBK deployments. This turned off the calculation of the METBK variables that use collocated VELPT data. A Helpdesk ticket (#14504) was submitted. The values for this calibration coefficient were changed back to 1 in Dec 2019, and the issue is resolved.

5. Pioneer Array Issues

5.1 Incorrect deployment location

The Deployment 8 location for CP02PMCO in Asset Management was more than 14 km away from the common platform deployment location. This was a result of a typo in the deployment sheet. A Helpdesk ticket (#13757) was submitted and the issue was resolved in December 2019.

5.2 Missing calibration coefficients

When originally downloaded for data review in June 2019 and March 2019, respectively, calibration values were missing for CP04OSSM-MFD35-02-PRESFC000 deployment 2 and CP04OSSM-SBD11-06-METBKA000 deployments 6 and 7. According to the provenance.json files downloaded with the data in April 2020, the PRESF calibration values are still missing but the METBK calibration coefficients have been added to the system.

5.3 Fluorometers on profilers

The fluorometers on Pioneer profilers have consistent issues with data quality during the second half of many deployments (Fig. 19). This should be investigated and resolved, or annotated to alert users of known issues with these instruments.
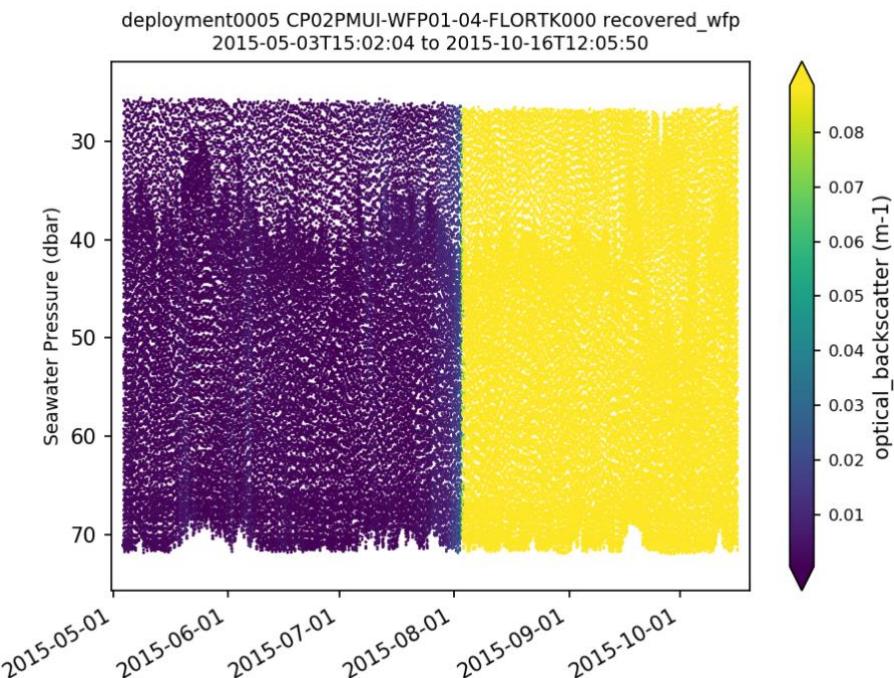
Figure 19. Example of data quality issue for fluorometers on Pioneer profilers.

6. Global Array Issues

6.1 Missing deployment depths

Deployments 4 and 5 of the Papa Moorings (GP02HYPM, GP03FLMA, GP03FLMB) are missing deployment depths in Asset Management.

6.2 Missing recovered data

The bottom portions of the Southern Ocean Flanking Moorings (GS03FLMA, GS03FLMB) were recovered in December 2018, but the recovered data are not available for download as of April 2020. These data need to be ingested or the datasets should be annotated to explain why the data are not available. Additionally, annotation IDs 378-379 still read "recovery options pending" and need to be updated.

GP03FLMB recovered_host CTDMO data (12 instruments) are not available for download for deployment 5. A Redmine ticket (#13628) was submitted in September 2018 and these data are still not available for download as of April 2020.

6.3 Global Hybrid Profiler data issues

An offset between recovered_wfp and telemetered data with the same timestamps is apparent for all deployments of all CTDs on the Global Hybrid Profilers. The values recorded at the first timestamp at the top/bottom of the profile are the same, but as the profiler moves down/up the offset between the values recorded for the same timestamp become increasingly offset. When the profiler reaches the end of the profile, either at its deepest or shallowest point, it appears to reset and the pattern repeats (Fig. 20). The issue seems to be related to how the data are interpolated by the OOI cyberinfrastructure. A Helpdesk ticket (#13743) was submitted and the issue is still under investigation as of April 2020
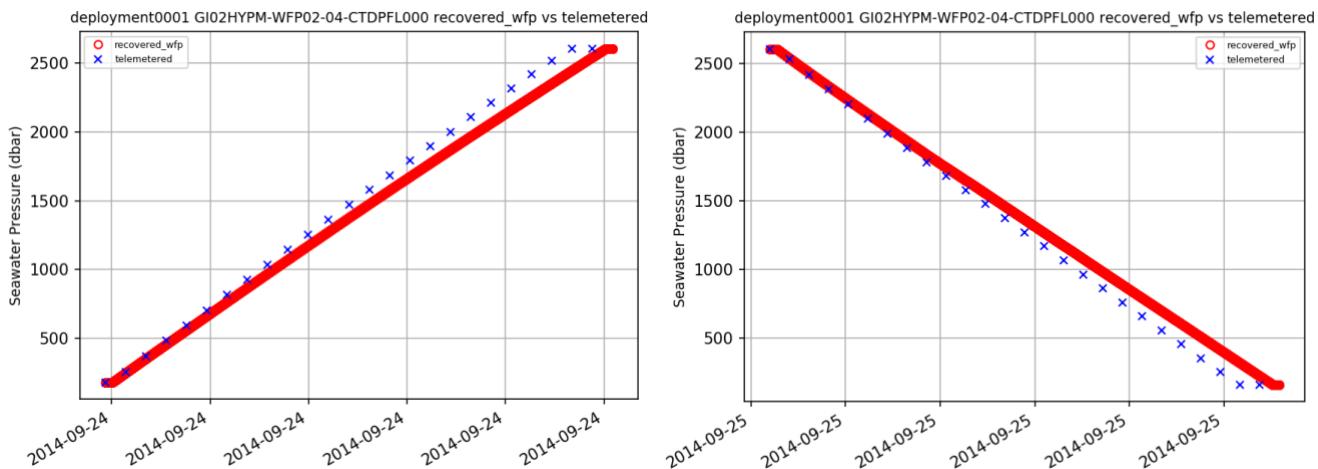
Figure 20. Recovered_wfp compared to telemetered data for one down- and up-cast showing the increasing offset in pressure over the length of the cast.

Dissolved oxygen values recorded by instruments on the deep Global Hybrid Profilers (2000+ dbar) shift substantially between some deployments. The issue was first noticed by a Subject Matter Expert in July 2018 for the Papa Hybrid Profiler (Fig. 21) and has also been observed on the Argentine Basin and Southern Ocean profilers. In addition, dissolved oxygen values at the bottom of the shallow profiler (WFP02) range are sometimes offset from the values recorded at the top of the deep profiler (WFP03) range (Fig. 22). These issues may be related to calibration issues. A Redmine ticket (#13505) was submitted in July 2018 and the issue is still under investigation as of April 2020.
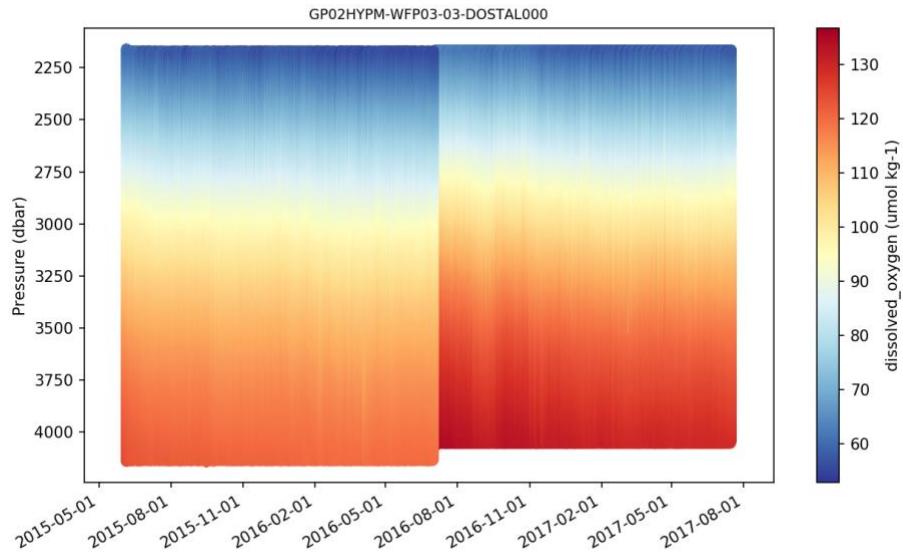
Figure 21. Unreasonable shift in Dissolved Oxygen between deployments 3 and 4 for GP02HYPM-WFP03-03-DOSTAL000
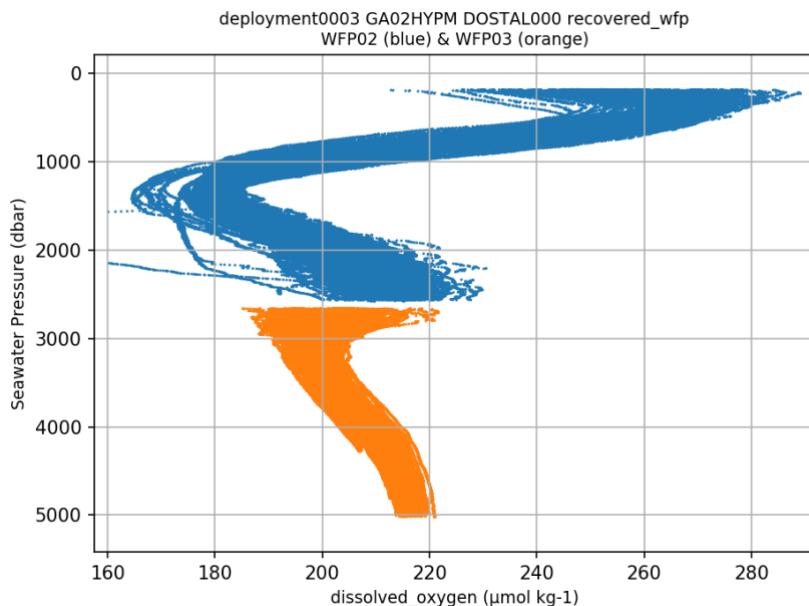
Figure 22. Example of the offset between values recorded by the Dissolved Oxygen sensors on the shallow (WFP02) and deep (WFP03) profilers.

6.4 Values between delivery methods do not match

CTD data from recovered and telemetered delivery methods from the same instrument do not match for several global CTD deployments (Table 15, Fig. 23). This could be related to calibration issues. A Redmine ticket (#12573) was submitted in August 2017 and the issue is still under investigation as of April 2020.

Table 15. Instrument deployments where data between different delivery methods do not match

Reference Designator	Deployment
GA03FLMB-RIM01-02-CTDMOG067	1
GI03FLMA-RIM01-02-CTDMOG042	2
GI03FLMA-RIM01-02-CTDMOG044	2
GI03FLMA-RIM01-02-CTDMOG048	4
GI01SUMO-RII11-02-CTDMOQ013	3
GP03FLMA-RIM01-02-CTDMOG041	2
GP03FLMA-RIM01-02-CTDMOG042	3
GP03FLMB-RIM01-02-CTDMOH070	3
GS01SUMO-RII11-02-CTDBPP033	2
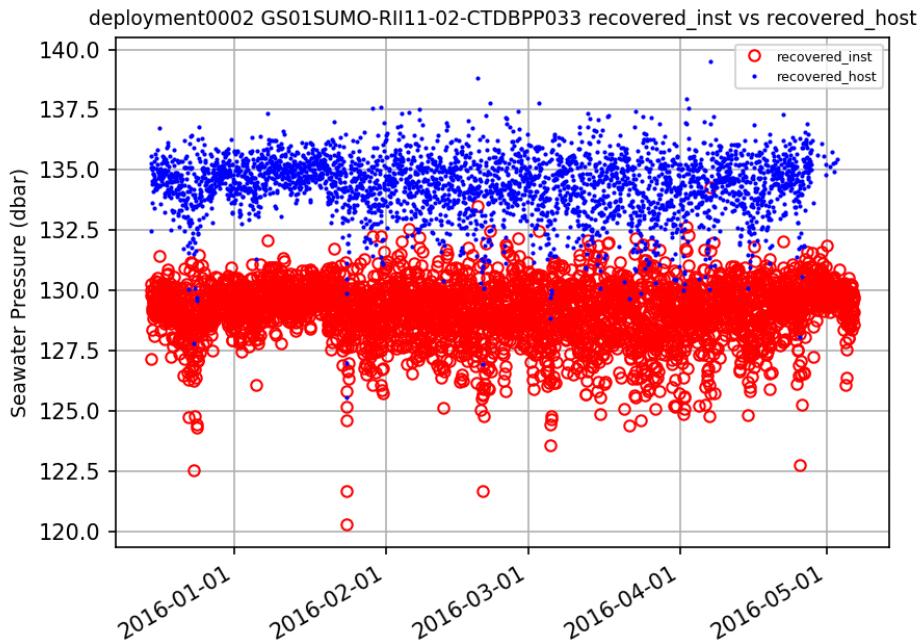
Figure 23. Example data from two different delivery methods (recovered_inst and recovered_host) from one instrument where the values are different for the same deployment.

7. Cabled Array Issues

7.1 Profilers

The Cabled Deep and Shallow Profilers are missing deployment depths in Asset Management. In addition, there have been several operational challenges regarding the Deep Profilers and the datasets should be annotated to alert users of the issues with these platforms.

7.2 CTD calibration offsets

Due to calibration errors for deployment 2 of several CTDs (Table 16), calibration offsets are being applied to the L1 Seawater Pressure data products for these instruments (see annotations on each dataset). The pressure data from these deployments, however, are still 5-120 dbar deeper compared to the pressure from the other deployments of each instrument (Fig. 24). The corrections should be reviewed to ensure they are being applied properly.

Table 16. Instruments for which deployment 2 Seawater Pressure data are suspect due to calibration errors, offsets should be reviewed.

Reference Designator
CE04OSPS-PC01B-4A-CTDPFA109
RS01SBPS-PC01A-4A-CTDPFA103
RS01SLBS-LJ01A-12-CTDPFB101
RS03AXBS-LJ03A-12-CTDPFB301
RS03AXPS-PC03A-4A-CTDPFA303

a.

b.
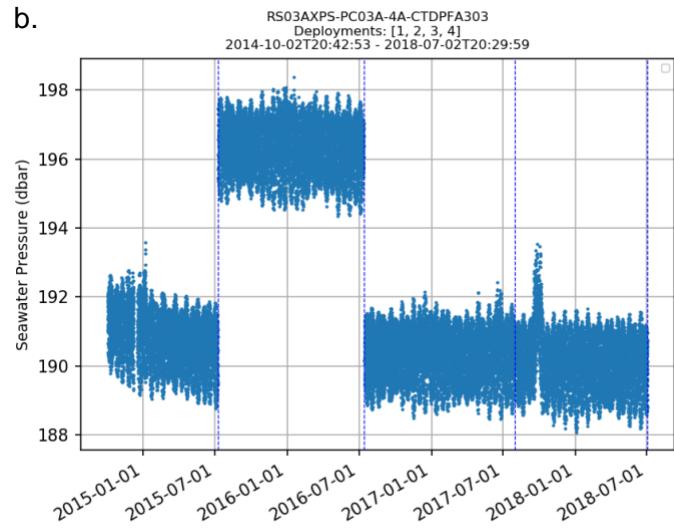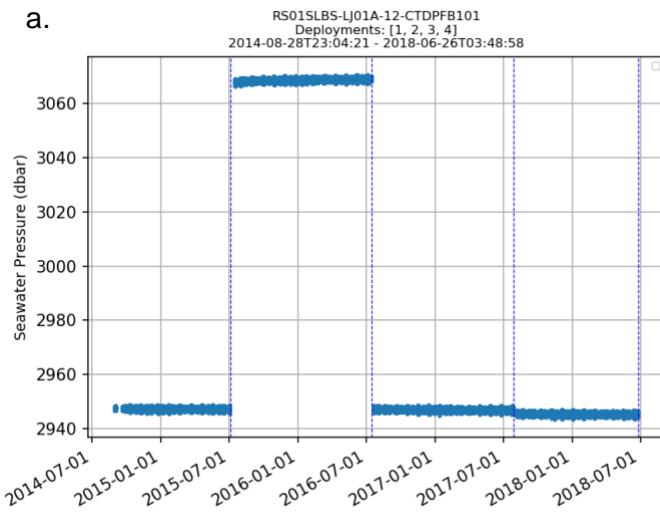
Figure 24. Deployment 2 with the applied pressure offset is still ~120 dbar and ~5 dbar deeper compared to the other three deployments of (a) RS01SLBS-LJ01A-12-CTDPFB101 and (b) RS03AXPS-PC03A-4A-CTDPFA303, respectively.

7.3 VADCP

The variables *bin_depths, pressure*, and *upward_seawater_velocity* are arrays of fill values for OOI 1.0 deployment NetCDF files.

7.4 D1000

Data from RS03INT1-MJ03C-07-D1000A301 are generally suspect. Data are often stuck at a specific value, or the sensors are not actually recording vent fluid temperatures (Fig. 25). This dataset should be annotated to alert users regarding issues with this instrument.
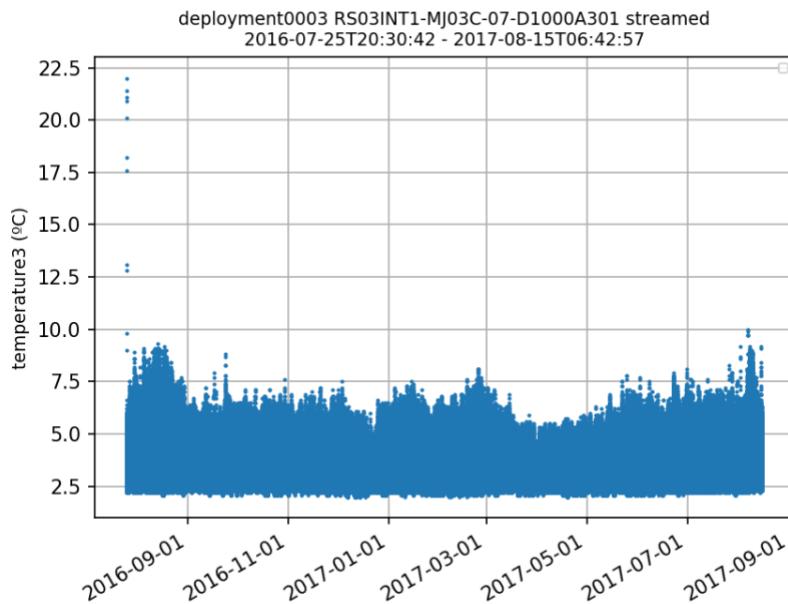
Figure 25. Values recorded by one of the thermistors on deployment 3 of RS03INT1-MJ03C-07-D1000A301 are not consistent with vent fluid temperatures.

7.5 DOSTA

Dissolved Oxygen data for the OOI 1.0 deployments are currently only available in the collocated CTD data stream (see Redmine #8662 for details). For data in OOI 2.0, however, the data stream containing dissolved oxygen has been separated from the CTD data stream and is now available via the DOSTA reference designator data streams. The OOI 1.0 dissolved oxygen data should similarly be ingested in DOSTA reference designator data streams so users can access the full dataset.

7.6 HPIES

The OOI is running an outdated version of THREDDS, and as such the NetCDF files for the echo_sounding HPIES data stream cannot be accessed via the OPENDAP link in the OOI THREDDs server. While the NetCDF files for this data stream can be downloaded and accessed locally, the OOI THREDDs server should be upgraded to have the capability to serve these data via the OPENDAP link. A Helpdesk ticket (#14559) was submitted and the OOI is researching the steps necessary to upgrade (as of April 2020).

7.7 TMPSF

The position of each thermistor in the 3D array is available in the provenance.json file as a calibration coefficient (e.g., CC_t15_position). While the information is available, it would be more convenient for the user if the positions are given as variable attributes in the NetCDF files.

8. Glider Issues

8.1 Glider latitude and longitude

The glider latitude and longitude variables were changed in September 2019 (see Redmine ticket #13182 for details), which required a purge and re-ingest of all glider data. Until the data re-ingest was completed, the latitude and longitude coordinates in the files contained arrays of NaNs and the descriptions of the variables were incorrect. A Helpdesk ticket (#14486) was submitted and all glider data were purged and re-ingested in December 2019. After the change, the majority (>95%) of the values for *sci_seawater_density* and *sci_abs_oxygen* were NaNs for several Endurance, Pioneer and Global gliders (Fig. 26). A new software package was released on Feb. 4, 2020 to fix the issue.
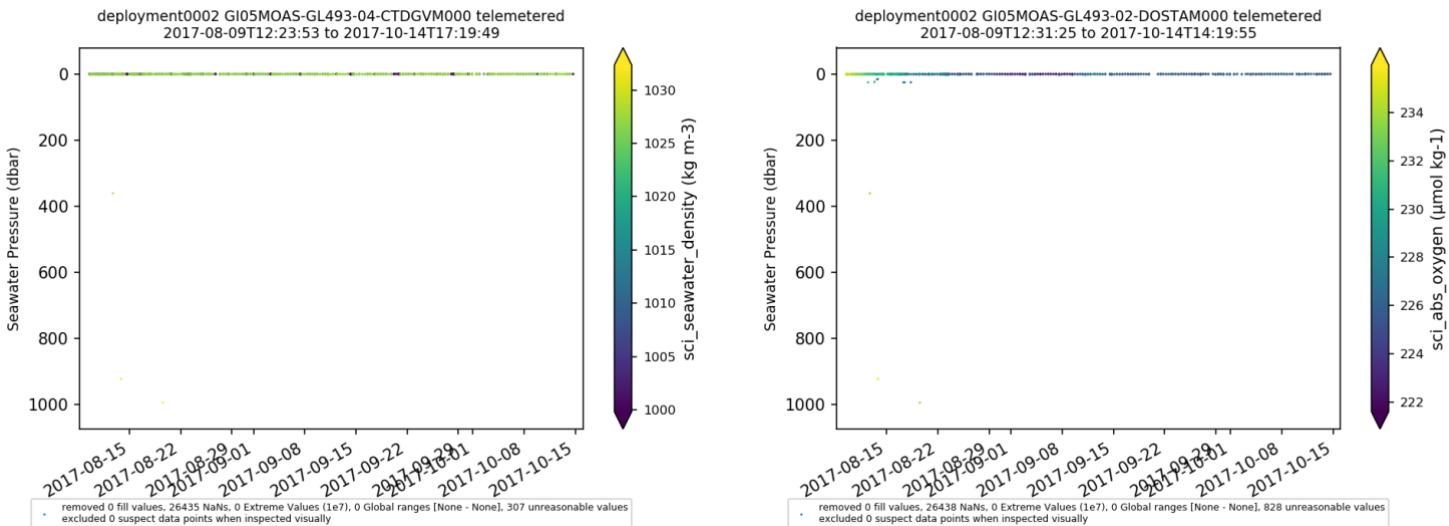
Figure 26. 98% of Seawater Density and Dissolved Oxygen values were NaNs for deployment 2 of GI05MOAS-GL493 after the re-assignment of the glider latitude and longitude variables.

As a result of the software release that fixed the NaN issue for Seawater Density and Dissolved Oxygen, recovered_host CTD data were then not available for download in NetCDF format. This was caused by a parameter renaming error prior to NetCDF generation from the previous software release. A Helpdesk ticket (#14626) was submitted and the issue was fixed on March 3, 2020.

After the software release that fixed the recovered glider NetCDF file generation, a substantial amount of recovered data was discovered to be missing from the downloaded NetCDF files that had been available prior to the renaming of the glider latitude and longitude variables. For example, there are now 103 days of missing CTD data, 91 days of missing DOSTA data, and 59 days of missing FLORD data for GP05MOAS-GL276 that were available when these datasets were last downloaded for analysis in April 2019 (Fig. 27). In addition, 30% of Seawater Density values are still NaNs, compared to 0% NaNs in the dataset previously downloaded in April 2019, and

Data Issues – Gliders

latitude and longitude are still missing from some glider NetCDF files. Helpdesk tickets 14654 and 14657 were submitted and the issues are still under investigation as of April 2020.

a.

b.
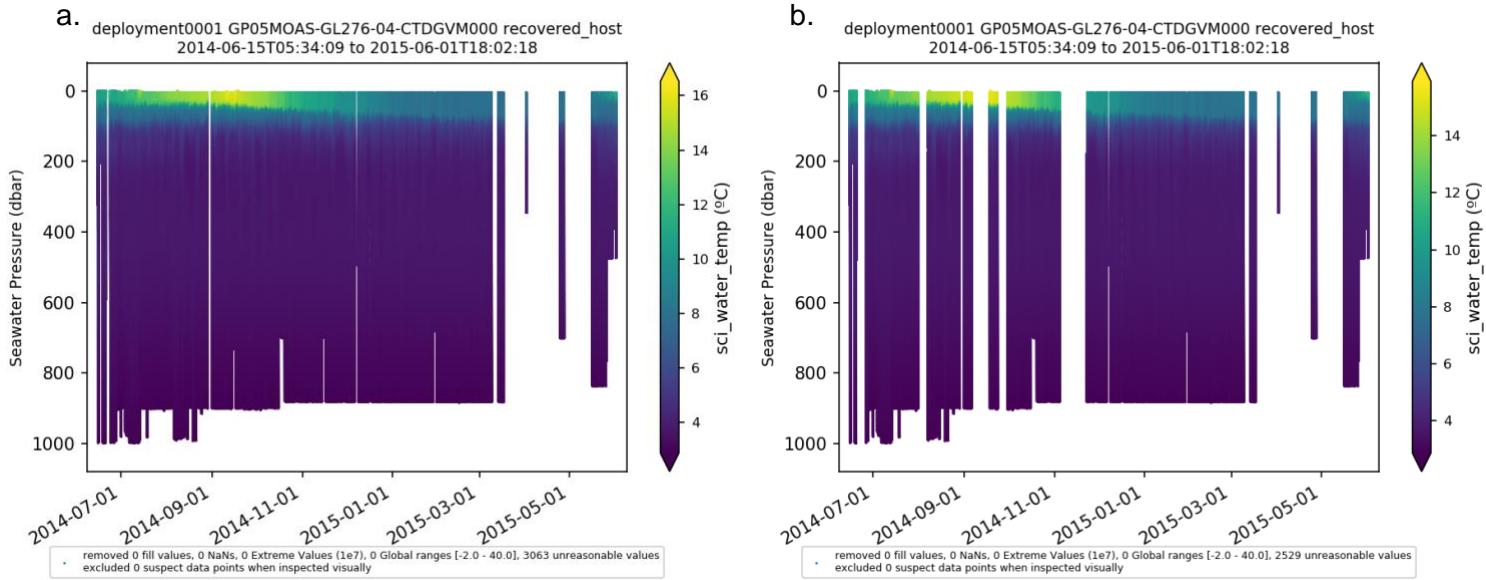
Figure 27. CTD data for GP05MOAS-GL276 downloaded on (a) April 12, 2019, and (b) March 3, 2020 after the renaming of the latitude and longitude variables and subsequent bug fixes.

8.2 Sampling pattern change

Annotation ID 1371 for CE05MOAS-GL319 explains an unintended sampling pattern change that was possibly caused by the ADCP. When the glider ADCP was turned off, the other sensors on the glider appeared to work properly again. CE05MOAS-GL383 deployments 2 and 4 appear to have the same issue and need annotations (Fig. 28).
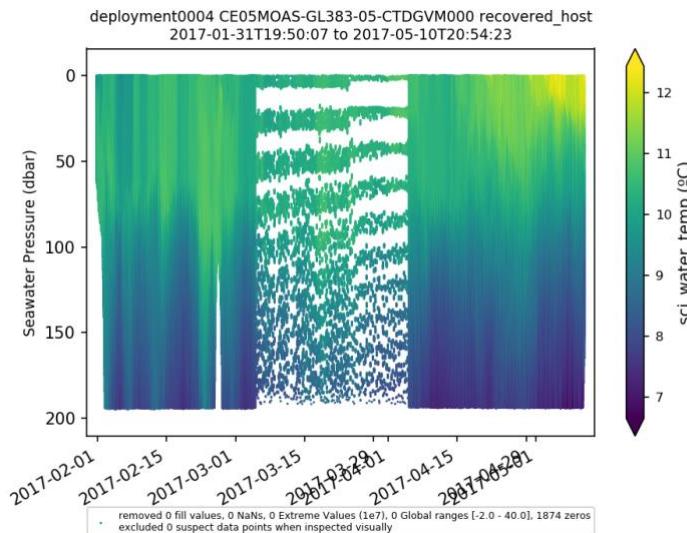
Figure 28. Apparent sampling pattern change in the middle of deployment 4 of CE05MOAS-GL383.

8.3 PARAD

For many glider deployments, the majority of PAR values are negative, and fail the Global Range QC test (Fig. 29). Also, *parad_m_par* values are orders of magnitude too high or low for several PARAD deployments, or the variable is an array of NaNs, which indicates that the *CC_bsipar_par_scaling* calibration factor is incorrect or missing (Table 17). The issue should be reviewed and annotated.
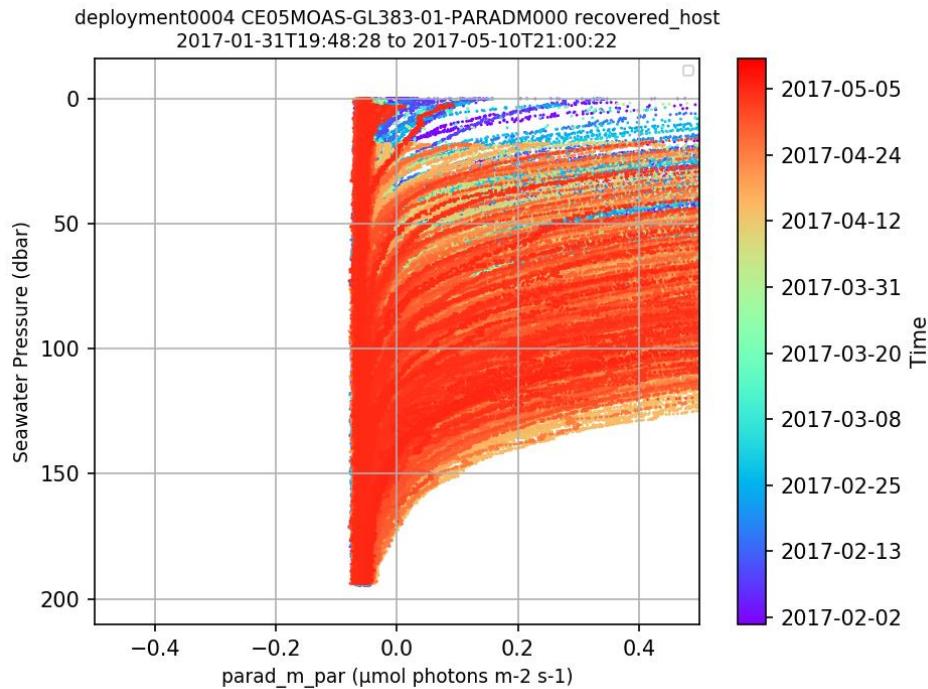
Figure 29. An example showing the majority (70%) of PAR data for one deployment are outside of global ranges (negative).

Data Issues – Gliders

Table 17. PARAD deployments for which *CC_bsipar_par_scaling* should be reviewed.

Reference Designator	Deployment(s)	Issue
CE05MOAS-GL326-01-PARADM000	3	*parad_m_par* values unreasonable
CP05MOAS-GL335-05-PARADM000	3	*parad_m_par* values unreasonable
CP05MOAS-GL335-05-PARADM000	5	missing *CC_bsipar_par_scaling*
CP05MOAS-GL336-05-PARADM000	3, 4, 5, 6	missing *CC_bsipar_par_scaling*
CP05MOAS-GL339-05-PARADM000	2, 3	*parad_m_par* values unreasonable
CP05MOAS-GL340-05-PARADM000	3, 6	*parad_m_par* values unreasonable
CP05MOAS-GL340-05-PARADM000	7	missing *CC_bsipar_par_scaling*
CP05MOAS-GL374-05-PARADM000	2	*parad_m_par* values unreasonable
CP05MOAS-GL376-05-PARADM000	3, 4	*parad_m_par* values unreasonable
CP05MOAS-GL379-05-PARADM000	3	*parad_m_par* values unreasonable
CP05MOAS-GL387-05-PARADM000	5, 6	*parad_m_par* values unreasonable
CP05MOAS-GL388-05-PARADM000	4, 5	*parad_m_par* values unreasonable
CP05MOAS-PG564-06-PARADM000	2	*parad_m_par* values unreasonable
CP05MOAS-PG583-06-PARADM000	1	*parad_m_par* values unreasonable
GA05MOAS-PG562-06-PARADM000	1	missing *CC_bsipar_par_scaling*
GA05MOAS-PG563-06-PARADM000	1	missing *CC_bsipar_par_scaling*
GA05MOAS-PG578-06-PARADM000	1	*parad_m_par* values unreasonable
GA05MOAS-PG580-06-PARADM000	1	*parad_m_par* values unreasonable
GP05MOAS-PG575-06-PARADM000	1	*parad_m_par* values unreasonable

8.4 FLORT and NUTNR

All FLORTO instruments are missing several calibration coefficients (Table 18), and CP05MOAS-GL336-02-FLORTM000 deployment 3 is missing *CC_scattering_angle*, *CC_measurement_wavelength*, *CC_depolarization_ratio* and *CC_angular_resolution*. These coefficients need to be added or flagged as missing and the datasets should be annotated.

Table 18. Missing calibration coefficients for all glider FLORTO instruments.

Calibration Coefficient Missing
CC_1_dark_counts_volume_scatter
CC_1_scale_factor_volume_scatter
CC_1_measurement_wavelength
CC_2_dark_counts_volume_scatter
CC_2_scale_factor_volume_scatter
CC_2_measurement_wavelength
CC_3_dark_counts_volume_scatter
CC_3_scale_factor_volume_scatter
CC_3_measurement_wavelength

Values for every FLORT parameter for CE05MOAS-GL386-02-FLORTM000 deployment 5 and CP05MOAS-GL340-02-FLORTM000 deployment 3 are close to zero for the entire deployment (Fig. 30). According to Redmine #12049, black electrical tape was still covering the sensor on the Endurance glider when it was recovered. Both of these datasets should be annotated.

Data Issues – Gliders
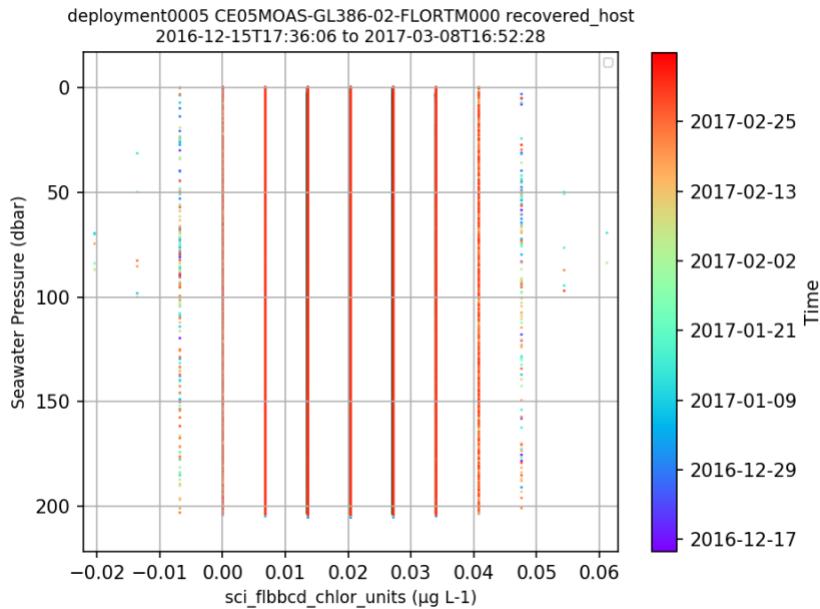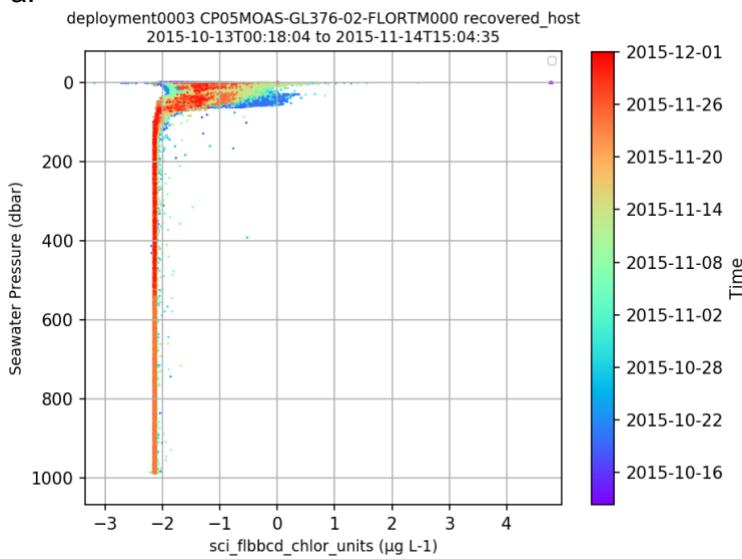
Figure 30. Chlorophyll-a data are close to zero for the entire deployment due to black electrical tape covering the sensor.

The majority (80-90%) of chlorophyll-a values are negative for CP05MOAS-GL376-02-FLORTM000 deployments 3-5, and nitrate values are mostly negative for CP05MOAS-PG564-05-NUTNRM000 deployment 2 (Fig. 31). These issues should be investigated and annotated.

a.

b.
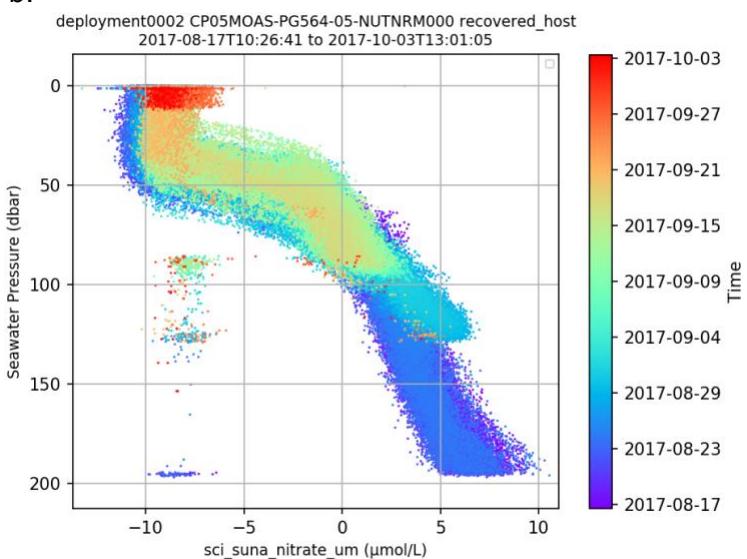
Figure 31. The majority of (a) chlorophyll-a values for deployment 3 of CP05MOAS-GL376 and (b) nitrate values for deployment 2 of CP05MOAS-PG564 are negative.

Supplementary Tables

Table S1.3a. Endurance data that need to be ingested or annotated to explain why they are not available for download.

Reference Designator	Delivery Method	Deployment	Notes
CE01ISSM-MFD35-01-VEL3DD000	recovered_inst	1, 2, 8, 9	
CE01ISSM-MFD35-05-PCO2WB000	recovered_inst	1-9	
CE01ISSM-RID16-03-CTDBPC000	recovered_inst	5, 7	Deployment 5: Instrument did not start logging - needs annotation
CE01ISSM-RID16-03-DOSTAD000	recovered_inst	5, 7	Deployment 5: Instrument did not start logging - needs annotation
CE01ISSM-RID16-04-VELPTA000	recovered_inst	6	
CE01ISSM-RID16-05-PCO2WB000	all	6	Instrument failed - needs annotation
CE01ISSM-RID16-05-PCO2WB000	recovered_inst	1-9	
CE01ISSM-RID16-06-PHSEND000	recovered_inst	4	
CE01ISSM-RID16-07-NUTNRB000	recovered_inst	1	
CE01ISSP-SP001-08-FLORTJ000	recovered	1-4, 6-7	
CE02SHSM-RID26-06-PHSEND000	recovered_inst	4	
CE02SHSM-SBD12-08-FDCHPA000	recovered_inst	6	
CE02SHSP-SP001-01-DOSTAJ000	all	1	
CE02SHSP-SP001-02-VELPTJ000	all	10	
CE02SHSP-SP001-04-OPTAAJ000	all	1	
CE02SHSP-SP001-06-SPKIRJ000	all	1	
CE02SHSP-SP001-07-FLORTJ000	recovered	1, 3, 4	
CE02SHSP-SP001-09-PARADJ000	recovered	1	
CE04OSSM-RID27-03-CTDBPC000	recovered_inst	6	
CE04OSSM-SBD12-04-PCO2AA000	recovered_host	5	SD card corrupted - needs annotation
CE04OSSM-SBD12-05-WAVSSA000	recovered_host	5	SD card corrupted - needs annotation
CE06ISSM-MFD35-01-VEL3DD000	recovered_inst	1-8	
CE06ISSM-MFD35-02-PRESFA000	recovered_inst	8	
CE06ISSM-MFD35-04-ADCPTM000	recovered_inst	2, 3, 8	
CE06ISSM-MFD35-05-PCO2WB000	recovered_inst	1-8	

Supplementary Tables

Reference Designator	Delivery Method	Deployment	Notes
CE06ISSM-MFD35-06-PHSEND000	recovered_inst	4, 8	
CE06ISSM-MFD37-03-CTDBPC000	recovered_inst	8	
CE06ISSM-MFD37-03-DOSTAD000	recovered_inst	8	
CE06ISSM-RID16-03-CTDBPC000	recovered_inst	8	
CE06ISSM-RID16-03-DOSTAD000	recovered_inst	8	
CE06ISSM-RID16-04-VELPTA000	all	8	
CE06ISSM-RID16-05-PCO2WB000	recovered_inst	1-8	
CE06ISSM-RID16-06-PHSEND000	recovered_inst	8	
CE06ISSM-RID16-07-NUTNRB000	recovered_inst	7	
CE06ISSM-SBD17-04-VELPTA000	recovered_inst	8	
CE06ISSM-SBD17-06-CTDBPC000	recovered_inst	1	
CE06ISSP-SP001	recovered	9	
CE07SHSM-MFD35-01-VEL3DD000	recovered_inst	2, 3-7	
CE07SHSM-MFD35-02-PRESFB000	recovered_inst	5, 7	
CE07SHSM-MFD35-04-ADCPTC000	recovered_inst	5, 7	
CE07SHSM-MFD35-05-PCO2WB000	recovered_inst	1-7	
CE07SHSM-MFD35-06-PHSEND000	all	1, 2	Data not expected - needs annotation
CE07SHSM-MFD35-06-PHSEND000	recovered_inst	5, 7	
CE07SHSM-MFD37-01-OPTAAD000	all	2	No data expected - needs annotation
CE07SHSM-MFD37-03-CTDBPC000	recovered_inst	7	
CE07SHSM-MFD37-03-DOSTAD000	recovered_inst	7	
CE07SHSM-RID26-01-ADCPTA000	recovered_inst	7	
CE07SHSM-RID26-04-VELPTA000	recovered_inst	7	
CE07SHSM-RID26-06-PHSEND000	recovered_inst	7	
CE07SHSM-RID26-07-NUTNRB000	recovered_inst	6	
CE07SHSM-RID27-03-CTDBPC000	recovered_inst	7	
CE07SHSM-SBD11-04-VELPTA000	recovered_inst	7	

Supplementary Tables

Reference Designator	Delivery Method	Deployment	Notes
CE07SHSM-SBD11-06-METBKA000	recovered_host	5	
CE07SHSM-SBD12-04-PCO2AA000	recovered_host	5	
CE07SHSM-SBD12-05-WAVSSA000	recovered_host	5	
CE09OSSM-MFD35-01-VEL3DD000	recovered_inst	1, 5-7	
CE09OSSM-MFD35-02-PRESFC000	recovered_inst	5, 7	
CE09OSSM-MFD35-04-ADCPSJ000	recovered_inst	5, 6, 7	Deployment 6: Instrument was damaged - needs annotation
CE09OSSM-MFD35-05-PCO2WB000	all	3	Instrument flooded - needs annotation
CE09OSSM-MFD35-05-PCO2WB000	recovered_inst	1-7	
CE09OSSM-MFD35-06-PHSEND000	recovered_inst, recovered_host	5	
CE09OSSM-MFD37-01-OPTAAC000	all	2	No data expected - needs annotation
CE09OSSM-MFD37-03-CTDBPE000	recovered_inst	7	
CE09OSSM-MFD37-03-DOSTAD000	all	2	Instrument failed - needs annotation.
CE09OSSM-MFD37-03-DOSTAD000	recovered_inst	7	
CE09OSSM-RID26-01-ADCPTC000	recovered_inst	7	
CE09OSSM-RID26-04-VELPTA000	recovered_inst	7	
CE09OSSM-RID26-06-PHSEND000	recovered_inst	7	
CE09OSSM-RID26-07-NUTNRB000	recovered_inst	6	
CE09OSSM-RID27-01-OPTAAD000	recovered_host	4	
CE09OSSM-RID27-03-CTDBPC000	recovered_inst	7	
CE09OSSM-SBD11-04-VELPTA000	recovered	7	
CE09OSSM-SBD11-06-METBKA000	recovered_host	7	
CE09OSSM-SBD12-04-PCO2AA000	recovered_host	7	
CE09OSSM-SBD12-05-WAVSSA000	recovered_host	7	

Supplementary Tables

Table S1.3b. Pioneer data that need to be ingested or annotated to explain why they are not available for download.

Reference Designator	Delivery Method	Deployment	Notes
CP01CNSM-SBD12-08-FDCHPA000	recovered_inst	4	
CP01CNSP-SP001-09-FLORTJ000	recovered	4	
CP04OSSM-SBD12-04-PCO2AA000	recovered_host	8	
CP04OSSM-SBD11-06-METBKA000	recovered_host	8	
CP04OSSM-RID26-08-SPKIRB000	recovered_host	8	
CP04OSSM-RID26-04-VELPTA000	recovered_inst	8	
CP04OSSM-RID27-03-CTDBPC000	recovered_inst	8	
CP04OSSM-RID27-04-DOSTAD000	recovered_host	8	
CP04OSSM-RID27-04-DOSTAD000	recovered_host	1	Raw data files are blank - needs annotation
CP04OSSM-MFD37-04-DOSTAD000	recovered_host	2	
CP04OSSM-SBD11-06-METBKA000	recovered_host	1	
CP01CNSM-RID26-06-PHSEND000	recovered_inst	7	
CP01CNSM-MFD35-06-PHSEND000	recovered_inst	6, 7	
CP04OSSM-RID26-06-PHSEND000	recovered_inst	1, 2	Data not expected - needs annotation
CP01CNSM-RID26-04-VELPTA000	recovered_inst	4, 5	
CP01CNSM-MFD35-04-VELPTA000	recovered_inst	3, 5	
CP02PMCI-WFP01	recovered	2	
CP03ISSM-MFD37-01-OPTAAD000	recovered_host	1	Data not expected - needs annotation
CP04OSSM-MFD37-01-OPTAAD000	recovered_host	2	
CP04OSSM-RID27-01-OPTAAD000	recovered_host	1	Raw data files are blank - needs annotation
CP03ISSP	recovered	3	Redmine 12790 - extract files missing
CP04OSSM-RID26-07-NUTNRB000	recovered_inst	1, 2	Deployment 1: Instrument failed - needs annotation
CP01CNSM-MFD35-05-PCO2WB000	recovered_inst	1-8	
CP03ISSM-MFD35-05-PCO2WB000	recovered_inst	1-7	
CP04OSSM-MFD35-05-PCO2WB000	recovered_inst	1-7	

Supplementary Tables

CP03ISSM-SBD12-04-PCO2AA000	all	1, 2	
CP03ISSM-MFD35-04-VELPTA000	recovered_inst	3, 4	
CP03ISSM-RID26-04-VELPTA000	recovered_inst	1, 4	
CP04OSSM-MFD35-04-VELPTB000	recovered_inst	3-5	
CP04OSSM-RID26-04-VELPTA000	recovered_inst	1, 3, 4, 5	Data not expected for 1 and 4 - needs annotation
CP03ISPM-RII01-02-ADCPTG010	recovered_inst	1	
CP02PMCO-RII01-02-ADCPTG010	recovered_inst	6	
CP02PMUO-RII01-02-ADCPSL010	recovered_inst	1	Raw data files are in the wrong folder on the raw data server
CP04OSPM-WFP01-01-VEL3DK000	recovered_wfp	4	
CP05MOAS-GL336	recovered_host	6	
CP05MOAS-GL340-01-ADCPAM000	recovered_host	4	
CP05MOAS-GL340-03-CTDGVM000	recovered_host	4	

Supplementary Tables

Table S1.3c. Global data that need to be ingested or annotated to explain why they are not available for download.

Reference Designator	Delivery Method	Deployment	Notes
GA01SUMO-RII11-02-ADCPSN010	recovered_inst	3	
GS03FLMA-RIM01-02-ADCPSL003	all	3	
GS03FLMB-RIM01-02-ADCPSL007	all	3	
GA01SUMO-RID16-05-PCO2WB000	recovered_inst	1-3	
GA01SUMO-RII11-02-PCO2WC051	recovered_inst	1-3	
GA01SUMO-RII11-02-PCO2WC052	recovered_inst	2-3	
GA01SUMO-RII11-02-PCO2WC053	recovered_inst	1-3	
GA01SUMO-RII11-02-PCO2WC052	recovered_host	1	
GA01SUMO-RID16-04-VELPTA000	recovered_inst	3	
GI03FLMB-RIM01-02-CTDMOH069	recovered_inst	4	Data not expected - needs annotation
GI01SUMO-RII11-02-CTDBPP031	recovered_inst	4	Data not expected - needs annotation
GI01SUMO-RII11-02-CTDBPP032	recovered_inst	4	Data not expected - needs annotation
GI01SUMO-RII11-02-CTDBPP033	recovered_inst	4	Data not expected - needs annotation
GI01SUMO-RID16-06-DOSTAD000	all	1	
GI01SUMO-RID16-05-PCO2WB000	recovered_inst	2-3	
GI01SUMO-RII11-02-PCO2WC051	recovered_inst	2	
GI01SUMO-RII11-02-PCO2WC052	recovered_inst	3	
GI01SUMO-RII11-02-PCO2WC053	recovered_inst	2-3	
GI01SUMO-SBD11-08-NUTNRB000	telemetered	4	
GI05MOAS-GL484	recovered_host	2	
GI05MOAS-GL493	recovered_host	2	There is no end date in asset management for this glider
GP03FLMA-RIM01-02-CTDMOG040	recovered_inst	2	
GP03FLMB CTDMOs	recovered_host	5	See Redmine 13628
GP03FLMB-RIM01-02-CTDMOG064	recovered_inst	2	
GP03FLMA-RIS01-04-PHSENF000	recovered_inst	4	

Supplementary Tables

GP03FLMB-RIS01-04-PHSENF000	recovered_inst	4	
GS01SUMO-RII11-02-CTDBPP031	all	2	
GS01SUMO-RII11-02-DOSTAD031	all	2	
GS01SUMO-RII11-02-FLORDG031	all	2	
GS03FLMA-RIM01-02-CTDMOG048	recovered	3	
GS03FLMA-RIM01-02-CTDMOH049	recovered	3	
GS03FLMA-RIM01-02-CTDMOH050	recovered	3	
GS03FLMA-RIM01-02-CTDMOH051	recovered	3	
GS03FLMB-RIM01-02-CTDMOG068	recovered	3	
GS03FLMB-RIM01-02-CTDMOH069	recovered	3	
GS03FLMB-RIM01-02-CTDMOH070	recovered	3	
GS03FLMB-RIM01-02-CTDMOH071	recovered	3	
GS01SUMO-RID16-05-PCO2WB000	recovered_inst	2	
GS01SUMO-RII11-02-PCO2WC051	recovered_inst	1-2	
GS01SUMO-RII11-02-PCO2WC052	recovered_inst	2	
GS01SUMO-RII11-02-PCO2WC053	recovered_inst	1-2	
GS05MOAS-GL485	recovered_host	1	Data not expected - needs annotation

Table S1.3d. Cabled data that need to be ingested or annotated to explain why they are not available for download.

Reference Designator	Deployment
CE04OSPS-PC01B-4B-PHSENA106	1, 4
CE04OSBP-LJ01C-08-OPTAAC104	1
CE04OSBP-LJ01C-09-PCO2WB104	2, 3
CE04OSPD-DP01B	3
CE04OSPD-DP01B-04-FLNTUA103	2
CE04OSPS-SF01B-3B-OPTAAD105	1
CE04OSPS-SF01B-3D-SPKIRA102	1
CE04OSPS-SF01B-3C-PARADA102	1, 2
RS03INT1-MJ03C-07-D1000A301	1
RS01SBPS-SF01A-4F-PCO2WA101	3
RS03AXPS-SF03A-3B-OPTAAD301	1

Table S1.5a. Endurance data are missing from the preferred stream for some portion of the deployment indicated (data are available in a non-preferred stream).

Reference Designator	Deployment
CE01ISSM-MFD37-03-CTDBPC000	7
CE01ISSM-MFD37-03-DOSTAD000	7
CE01ISSM-RID16-03-CTDBPC000	6
CE01ISSM-RID16-03-DOSTAD000	2, 6
CE01ISSM-SBD17-06-CTDBPC000	2
CE06ISSM-MFD37-03-CTDBPC000	5, 6
CE06ISSM-MFD37-03-DOSTAD000	5, 6
CE06ISSM-RID16-03-CTDBPC000	1, 4, 6
CE06ISSM-RID16-03-DOSTAD000	1, 4, 6
CE09OSSM-RID26-06-PHSEND000	6
CE02SHSM-SBD11-06-METBKA000	6
CE07SHSM-SBD11-06-METBKA000	4
CE01ISSM-RID16-06-PHSEND000	2, 5, 7
CE02SHSM-RID26-06-PHSEND000	2, 6, 7
CE07SHSM-RID26-06-PHSEND000	2
CE04OSSM-RID26-06-PHSEND000	5, 6
CE02SHSM-SBD12-05-WAVSSA000	4
CE09OSSM-SBD12-05-WAVSSA000	4
CE07SHSM-RID26-04-VELPTA000	1
CE06ISSM-MFD35-04-ADCPTM000	5

Table S1.5b. Pioneer data are missing from the preferred stream for some portion of the deployment indicated (data are available in a non-preferred stream).

Reference Designator	Deployment
CP03ISSM-MFD35-06-PHSEND000	6
CP03ISSM-MFD37-04-DOSTAD000	4
CP03ISSM-RID27-04-DOSTAD000	5
CP01CNSM-SBD11-06-METBKA000	7
CP04OSSM-SBD11-06-METBKA000	3
CP02PMCI-WFP01-01-VEL3DK000	1, 3, 4
CP02PMCO-WFP01-01-VEL3DK000	1, 2, 3
CP04OSSM-RID26-04-VELPTA000	3
CP02PMUO-WFP01-01-VEL3DK000	5
CP04OSPM-WFP01-01-VEL3DK000	1, 8

Table S1.5c. Global data are missing from the preferred stream for some portion of the deployment indicated (data are available in a non-preferred stream).

Reference Designator	Deployment
GA01SUMO-SBD12-05-WAVSSA000	2
GA01SUMO-RII11-02-CTDMOQ014	3
GA01SUMO-RII11-02-PHSENE042	2
GA01SUMO-SBD12-06-METBKA000	2
GA01SUMO-SBD12-04-PCO2AA000	2
GI01SUMO-RID16-03-CTDBPF000	3
GI01SUMO-SBD11-04-DOSTAD000	3
GI01SUMO-RID16-06-DOSTAD000	3
GI01SUMO-SBD12-02-FLORTD000	3
GI01SUMO-RID16-02-FLORTD000	3
GI01SUMO-SBD11-06-METBKA000	3
GI01SUMO-SBD12-06-METBKA000	3
GI01SUMO-RII11-02-PCO2WC051	2
GI01SUMO-RII11-02-PCO2WC052	2
GI01SUMO-RID16-04-VELPTA000	1
GP03FLMA-RIM01-02-CTDMOG040	5
GP03FLMA-RIM01-02-CTDMOG041	2
GP03FLMA-RIM01-02-CTDMOG044	2
GP03FLMB-RIM01-02-CTDMOG062	2
GP03FLMB-RIS01-03-DOSTAD000	2
GS01SUMO-RID16-06-DOSTAD000	2
GS01SUMO-SBD11-04-DOSTAD000	1
GS01SUMO-SBD11-06-METBKA000	2
GS01SUMO-RII11-02-PCO2WC051	2
GS01SUMO-RII11-02-PCO2WC052	2
GS01SUMO-RII11-02-PCO2WC053	2