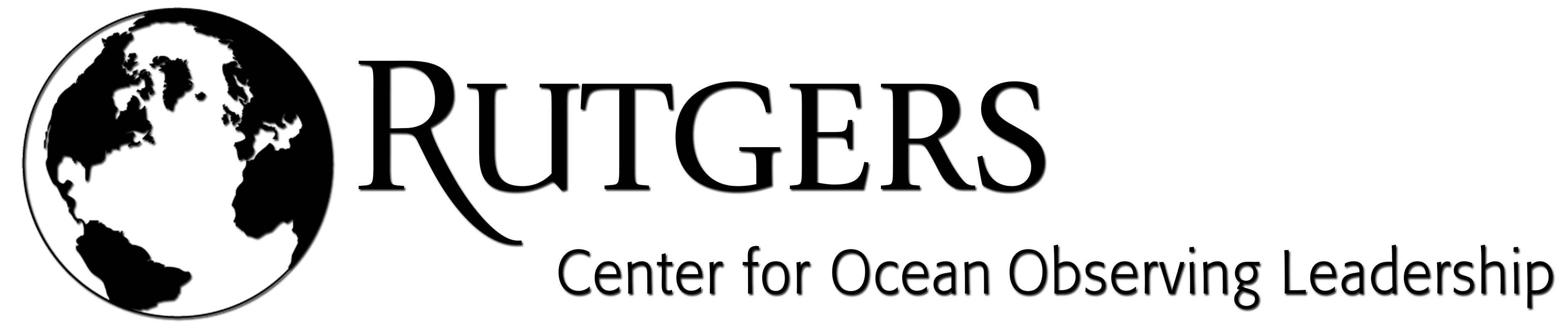
****

***OOI Synthesis & Education:***

***Educational Support and Synthesis Based on the Initial***

***Phase of the Ocean Observatories Initiative.***

***Consortium for Ocean Leadership***

***National Science Foundation Grant No. OCE-1841799***

**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**

**Overview………………………………………………………………………………………………………… 2**

**Data Evaluation Process………………………...……………………………………………………….. 3**

**Summary of Data Issues……………………………….………………………………………………….. 5**

1.0 General Data Issues Affecting All Arrays……………………………..………….……….. 5

2.0 Instrument-specific Issues………………………………………………………………………..

3.0 Endurance Array Issues………………………………………………………………………

4.0 Pioneer & Global Array Issues………………………………………………………………

5.0 Pioneer Array Issues…………………………………………………………………………

6.0 Global Array Issues…………………………………………………………………………..

7.0 Cabled Array Issues…………………………………………………………………………..

**OOI 1.0 Data Evaluation Summary Report**

**Overview**

The Ocean Observatories Initiative (OOI: <https://oceanobservatories.org/>) 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 the OOI has become operational in 2016, data from two coastal arrays (CP, CE), one cabled array (RS), and four global observatories (GA, GI, GP, GS) are being made available through the OOI data portal (OOINET: <https://ooinet.oceanobservatories.org/>). 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 conducted by Rutgers University and led by the Consortium for Ocean Leadership, the Data Team at Rutgers reviewed and evaluated the official datasets collected by the OOI from July 2013 to September 2018 (OOI 1.0). The goal was to 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. Another goal was to identify and collate OOI datasets that could be used by educators in the classroom. In this report, only tasks associated with the data evaluation process will be addressed.

*Instruments for Review*

There are 1327 instruments (defined as individual reference designators) in the OOI system from July 2013 to September 2018. For the review process we selected instruments that contain science parameters and that have data downloadable via OOINET. Each instrument has multiple deployments, some have multiple data streams, and uncabled instruments can have multiple data delivery methods: via telemetry and through various recovery methods. To limit the number of datasets to a manageable size, we selected one data delivery method 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).

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

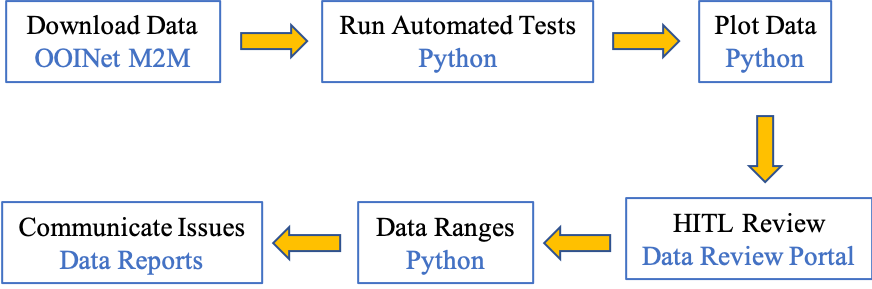
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Review Category | CE | CP | GA | GI | GP | GS | RS | **Total** |
| Science Instruments | 247 | 186 | 128 | 128 | 85 | 116 | 67 | **957** |
| Instrument-deployment-streams | 1829 | 1184 | 302 | 426 | 253 | 242 | 216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

\* Instruments not for review:

* Cameras
* AUVs
* Instrument not delivering data via uFrame (e.g. ZPLSC)
* Engineering instruments
* Sensor deployments recovered after 30 September 2018
* Instruments with no data before 30 September 2018

**Data Evaluation Process**

For each instrument flagged for review, we downloaded data for the deployments within the OOI 1.0 timeframe through OOINET via the Machine-to-Machine interface using a Python toolbox1. We then analyzed the downloaded files using an automated set of tools2 to test the basic components of each dataset (Table 2). The output of the automated tests were imported into the Data Review Portal3, and the data were plotted4 and stored in a public location5 to aid in the Human-In-The-Loop review. We then reviewed all relevant information from the automated tests, plots, and system annotations, and entered notes in the Data Review Portal regarding data issues that weren’t already explained in system annotations. Once the review was complete, final data ranges were calculated for each applicable dataset, excluding data outside of global ranges and suspect data6. These data ranges can be used to inform local ranges, as these ranges are largely absent from the OOI system. Finally, the review notes entered in the Data Review Portal were sent to COL as a monthly data report for further distribution to the OOI system and operator teams7.



*Links and repositories*

1Data download toolbox: <https://github.com/ooi-data-lab/data-download>

2Automated data analysis: <https://github.com/ooi-data-lab/data-review-tools/tree/master/data_review>

3Data Review Portal: <https://datareview.marine.rutgers.edu/>

4Plotting toolbox: <https://github.com/ooi-data-lab/data-review-tools/tree/master/plotting>

5Plot repository: <https://marine.rutgers.edu/cool/ooi/data-eval/data_review/>

6Final data ranges: <https://github.com/ooi-data-lab/data-review-tools/tree/master/data_review/data_ranges>

7Data reports: <https://github.com/ooi-data-lab/data-review-tools/tree/master/data_review/review_reports>

Table 2. Automated Data Quality Checks.

|  |  |
| --- | --- |
| 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) |

**Summary of Data Issues**

This summary was collated using Helpdesk tickets submitted via Redmine, as well as notes recorded in the Rutgers University Data Review Database (<https://datareview.marine.rutgers.edu/>). The notes were exported and shared with the Marine Implementing Organizations monthly, and are stored in the data review section of the OOI Data Labs Github repo (<https://github.com/ooi-data-lab/data-review-tools>).

**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. Large gaps in the data are often not annotated or explained. 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 |

Some annotations that are currently in the system should be reviewed and updated. There are several examples of duplicated annotations that should be deleted (e.g. RS01SBPS-SF01A-4A-NUTNRA101 annotation ID 1370 duplicates 572), and annotations from several years ago that indicate an investigation is pending (CE04OSPS-SF01B-2A-CTDPFA107), 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. Deviations from planned sampling rates should also be annotated (e.g. CP01CNSM-RID27-03-CTDBPC000 deployments 5-7).

During an OOI Ocean Data Labs workshop in March 2019, data requests were taking an exceedingly long time to fulfill. We submitted a Redmine ticket (ticket # 14236) on the first day of the workshop, when participants were mainly planning their projects and not actively trying to download data. We received a response that the systems team was being alerted, and received no further communication from the OOI team until 2 days later, which was after the workshop participants had attempted to download data. There was no explanation of the issue, just that they had also seen the delay and it “cleared up”. The workshop participants struggled to download data throughout the workshop, and we had no explanation or communication from the OOI regarding the nature of the issue or a timeline for resolution.

*1.2 Missing data*

Numerous instrument deployments for each array are not available for download from the OOI cyberinfrastructure (Table 4) are not annotated to explain why the data aren’t available. These data either need to be ingested or annotated to explain why they aren’t available for download (full list: **Tables S1-S4**).

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

|  |  |  |
| --- | --- | --- |
| Array | # Instruments | Details |
| Endurance | 78 | Table S1 |
| Pioneer | 42 | Table S2 |
| Global | 49 | Table S3 |
| Cabled | 15 | Table S4 |

*1.3 NetCDF file coordinates missing or incorrectly assigned*

An instrument’s location in three-dimensional space (latitude, longitude, and depth) is critical for data use and interpretation, and should be provided as coordinates in every NetCDF file generated by the OOI system. NetCDF files are currently missing one or all of these coordinates for the following instrument types (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 |

In some instances, the pressure coordinate isn’t included in NetCDF files for deployments where the collocated CTD data aren’t 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 isn’t available, the array should be filled with NaNs or fill values.

To exacerbate the problem, the system currently only pulls collocated data for the same delivery method when calculating L2 data products – for most DOSTAs and FLORTs, the delivery method is recovered\_host. If the collocated recovered\_host CTD data aren’t 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 the preferred delivery method isn’t available.

*1.4 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 downloaded directly from the instrument upon recovery. When this delivery method isn’t available (e.g. the instrument doesn’t 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 aren’t available in the preferred recovered\_inst data stream (full list: **Tables S5-S7**).

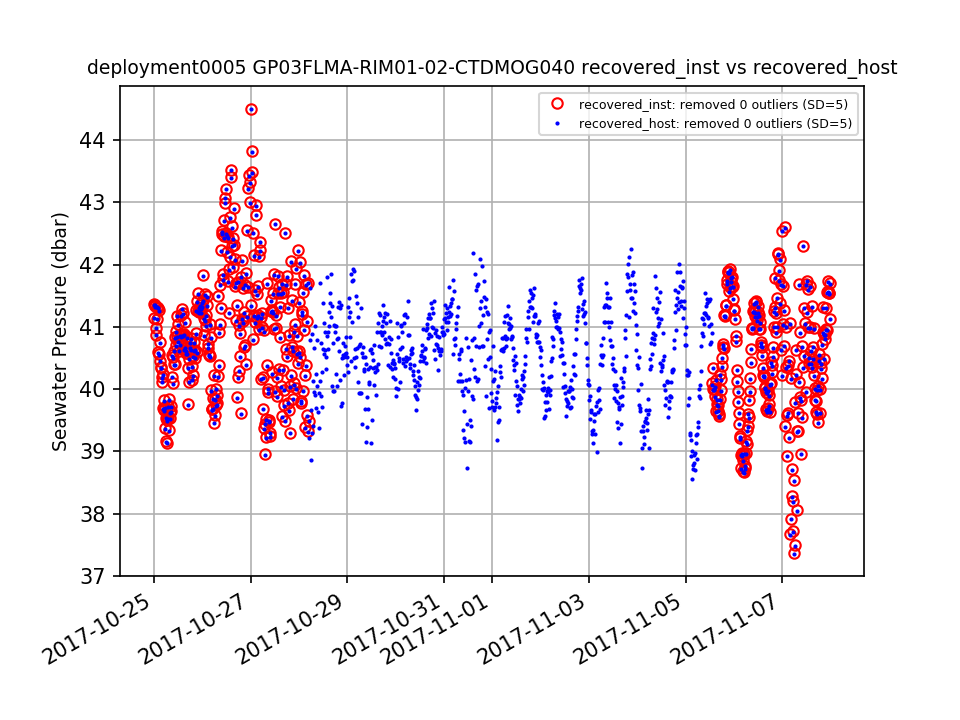


Figure 1. Example from a Global dataset where data are available in a “non-preferred” data stream (recovered\_host – blue dots), and not available in the “preferred” data stream (recovered\_inst – red circles) for a portion of the deployment.

*1.5 Incomplete and missing shipboard data*

Supplemental data collected on OOI maintenance cruises (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, etc. We attempted to do comparisons of data collected from OOI platforms with shipboard CTD casts, 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 we were able to do mostly with Pioneer and Global data, differences between the platforms and the shipboard CTD casts were often substantial, and 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 need to be uploaded.

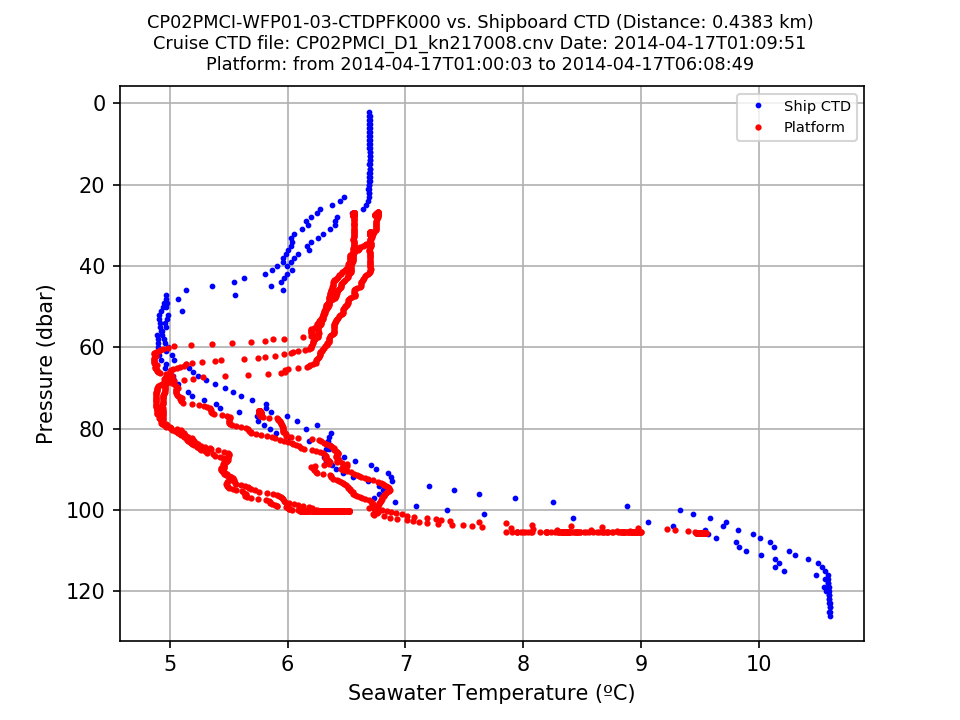


Figure 2. Example comparison of CTD data from deployment 1 of the Pioneer Central Inshore Profiler Mooring with a shipboard CTD cast done at the time of deployment.

*1.6 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. We identified several instruments and parameters that are either missing global ranges (Table 6), or the global ranges look incorrect and need 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. Instruments and parameters with 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 |

**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 to the upper detection limit of the instrument and 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, and if this is a known issue these datasets should be annotated to inform users of any issues (Table 8, Fig. 3).

Table 8. Instrument deployments showing a repetitive pattern of *total\_volume\_scattering\_coefficient* increasing to the upper detection limit and dropping to zero.

|  |  |
| --- | --- |
| Reference Designator | Deployment(s) |
| CE06ISSM-RID16-02-FLORTD000 | 1 |
| CE01ISSM-RID16-02-FLORTD000 | 2, 7 |
| CE02SHSM-RID27-02-FLORTD000 | 6 |
| CE04OSSM-RID27-02-FLORTD000 | 4 |

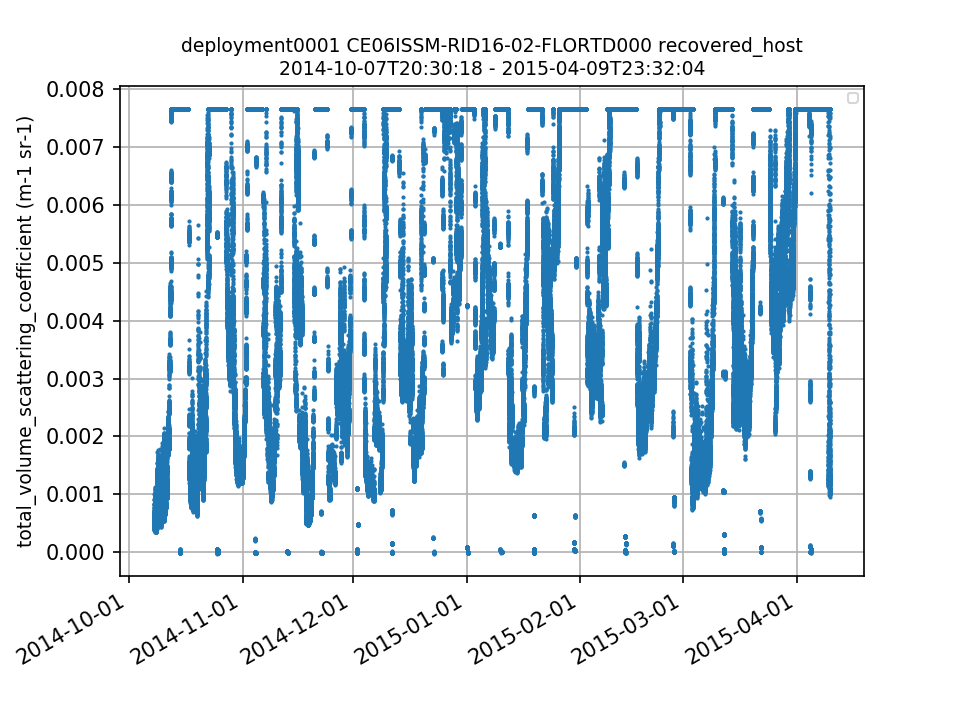


Figure 3. Timeseries plot of CE06ISSM-RID16-02-FLORTD000 showing a repetitive pattern of *total\_volume\_scattering\_coefficient* increasing to the upper detection limit and dropping to zero.

*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 corrections for these datasets should be provided to users.

Table 9. METBK instrument deployments where *shortwave\_irradiance* and *met\_netsirr* minimum nighttime values jump unreasonably.

|  |  |
| --- | --- |
| 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 |

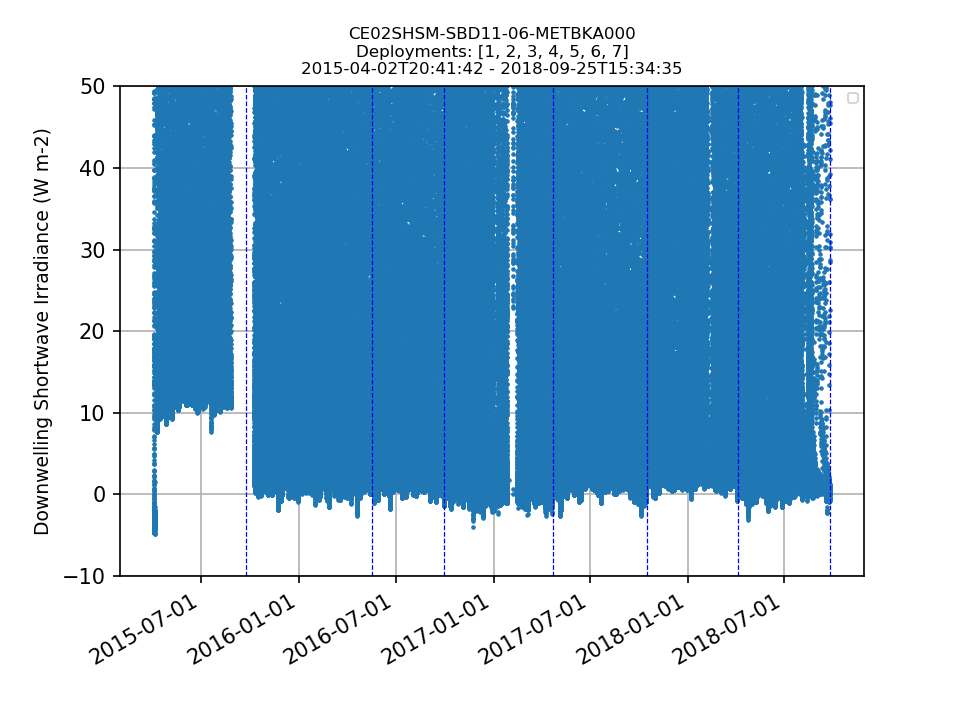


Figure 4. Example showing the jump in minimum values for Downwelling Shortwave Irradiance (*shortwave\_irradiance*) at the beginning of deployment 1 for CE02SHSM-SBD11-06-METBKA000.

For several METBK deployments where no CT data are expected (e.g. the CT sensor was not connected correctly), *sea\_surface\_temperature* values are -5, *sea\_surface\_conductivity* values are 0.0, and *met\_salsurf* values are 0.0 (Table 10, Fig. 5), when 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 instrument deployment below compared to the other deployments, 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 (Fig. 6).

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 |

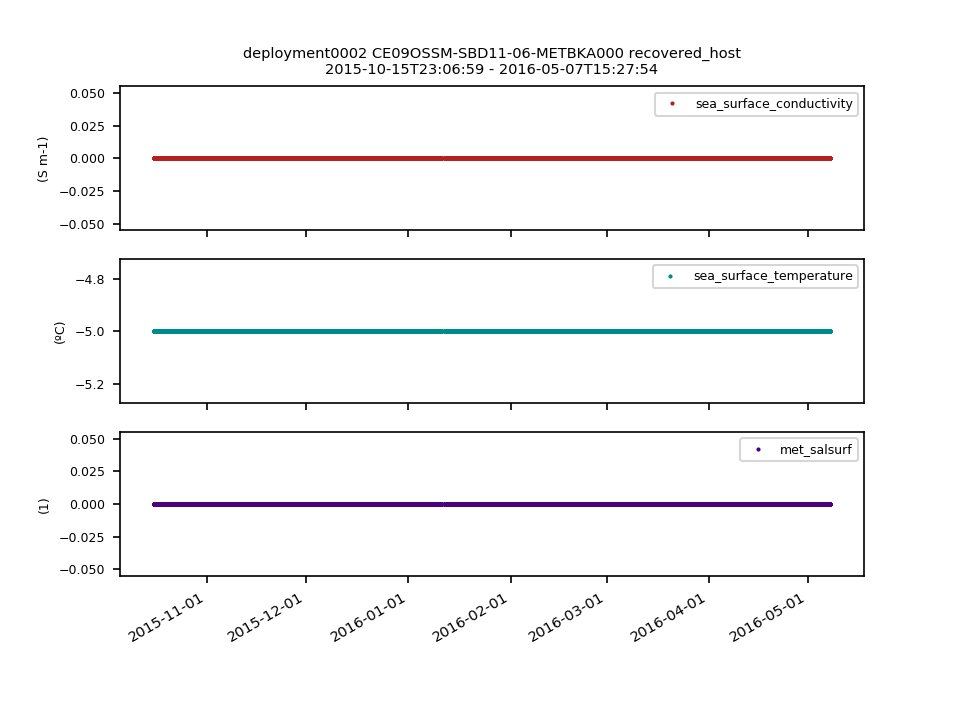


Figure 5. Deployment 2 of CE09OSSM-SBD11-06-METBKA000 where *sea\_surface\_temperature* = -5, *sea\_surface\_conductivity* = 0.0 and *met\_salsurf* = 0.0.

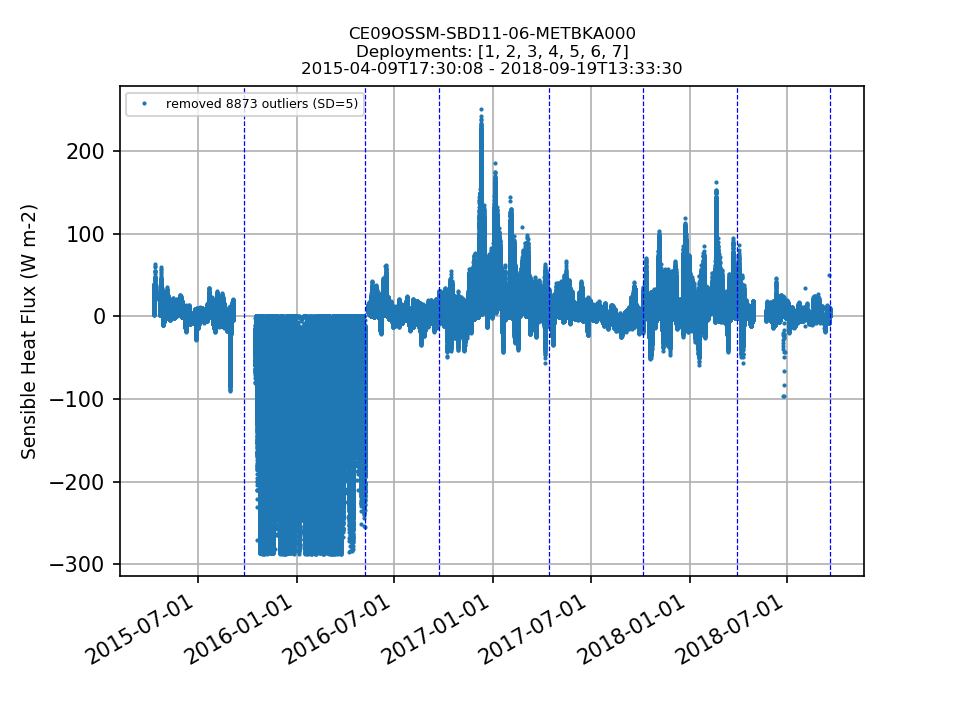


Figure 6. Sensible Heat Flux (*met\_sensflx\_minute*) for deployment 2 of CE09OSSM-SBD11-06-METBKA000 is substantially different compared to the other deployments of this instrument.

*2.4 NUTNR*

The annotation on all NUTNR instruments that were switched to SUNA 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. According to Roesler and Barnard1, ”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”.

The wavelength variables (*wavelength\_a* and *wavelength\_c*) are occasionally arrays of fill values. (e.g. CP01CNSM-RID27-01-OPTAAD000 deployments 7-8, CE01ISSM-MFD37-01-OPTAAD000 deployment 1). These variables are crucial for data analysis and interpretation.

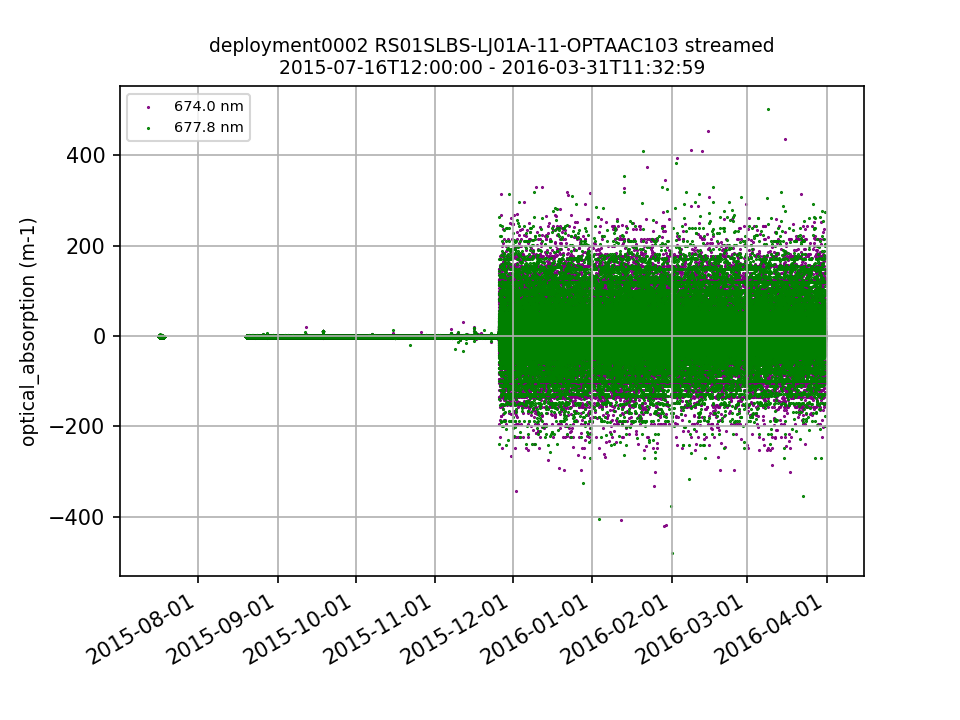


Figure 7. Optical Absorption values degrade over time for RS01SLBS-LJ01A-11-OPTAAC103.

1Roesler 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.

*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.

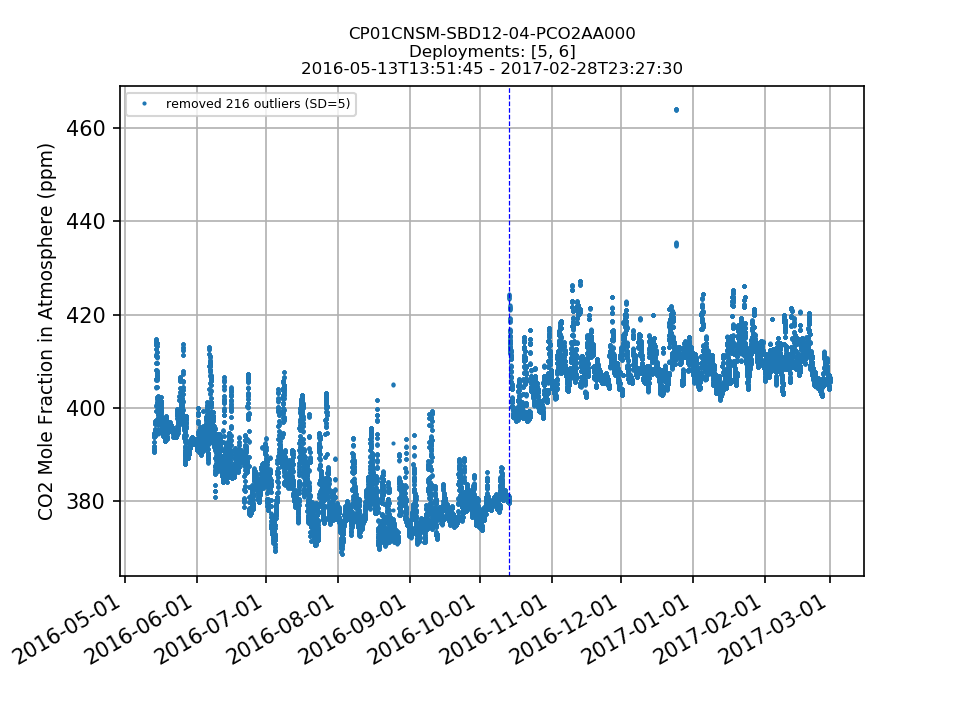


Figure 8. Values for CO2 Mole Fraction in Atmosphere from CP01CNSM-SBD12-04-PCO2AA000 are approximately 20 ppm lower at the end of deployment 5 compared to the values at the beginning of deployment 6.

**Redmine 14529:** Recovered data were not available for download for four deployments of Pioneer and Endurance PCO2A data. We were attempting to create Jupyter Notebook tutorials for educators using PCO2A data, so we submitted a Helpdesk ticket regarding the missing data. The Marine Implementing Organizations ingested these data into the OOI cyberinfrastructure or annotated the datasets to explain why the data aren’t available for download in November 2019 and the issue is resolved.

**Redmine 14531:** CO2 Flux (*pco2\_co2flux*) currently isn’t being provided by the system for Pioneer and Global datasets (see Redmine ticket #14527 for details). In order to create pCO2 air-sea flux tutorials for educators, we calculated this variable externally using the *pco2\_co2flux* function from the OOI ion\_functions. When comparing our results to those provided by the OOI system for Endurance datasets, we observed that the OOI cyberinfrastructure was inverting the flux direction (Fig. 9). The MIOs determined that the inputs to the function were reversed for *pco2a* and *pco2w*, and the issue was fixed in December 2019.

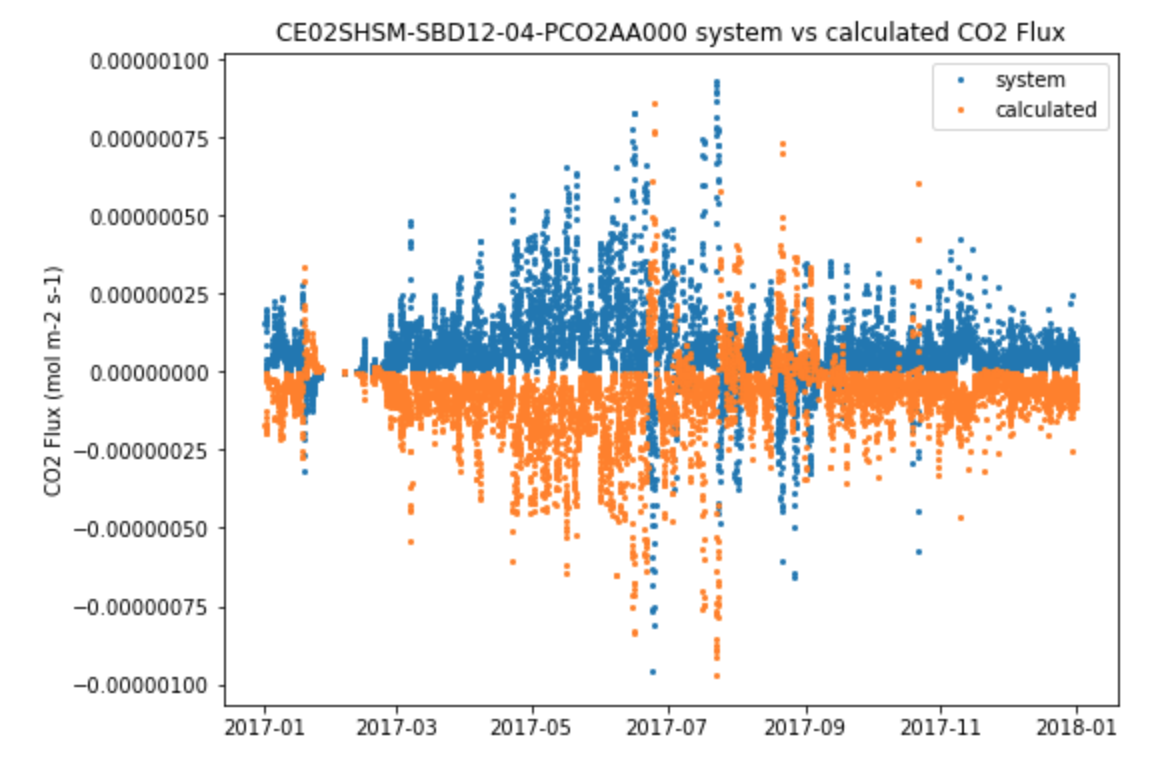


Figure 9. CO2 Flux (*pco2\_co2flux*) provided by the OOI cyberinfrastructure (blue) compared to CO2 Flux calculated externally (orange) before the resolution in December 2019.

*2.7 PCO2W*

Recovered\_inst data are not available for download for any uncabled instrument. According to the ingest .csvs, there are raw recovered\_inst data files available – these data should be ingested.

At the time of the analysis, 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 on Nov 8, 2019 to be 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’s currently an index from 0-1023 and is unclear to what this is referring.

*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).

**

Figure 10. For deployment 1 of GI01SUMO-RID16-08-SPKIRB000, 40% of the values for every wavelength are slightly negative and fail the Global Range QC test.

*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.

*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 11) and are not annotated. This should be investigated and the datasets should be annotated if there was a configuration mistake.

Table 11. 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 |

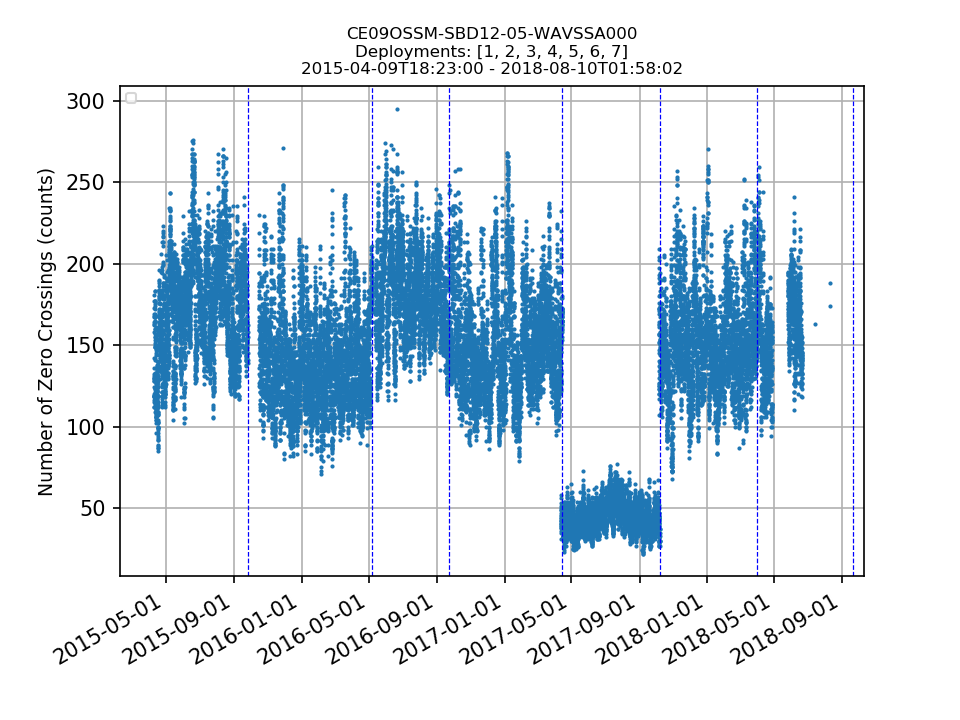


Figure 11. The Number of Zero Crossings for deployment 5 of CE09OSSM-SBD12-05-WAVSSA000 are substantially lower compared to all other deployments, suggesting a configuration error.

**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.

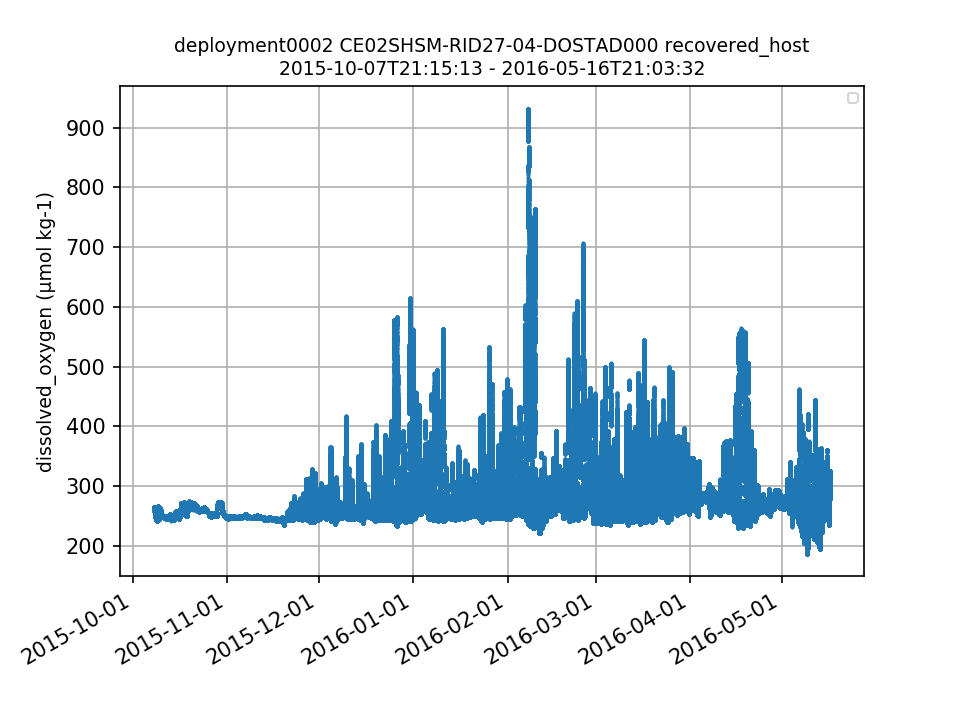
****

Figure 12. Example of unreasonable Dissolved Oxygen data, likely due to biofouling.

For several instrument turnovers, Dissolved Oxygen values at the end of one deployment don’t 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 these datasets should be annotated if there is a known issue.

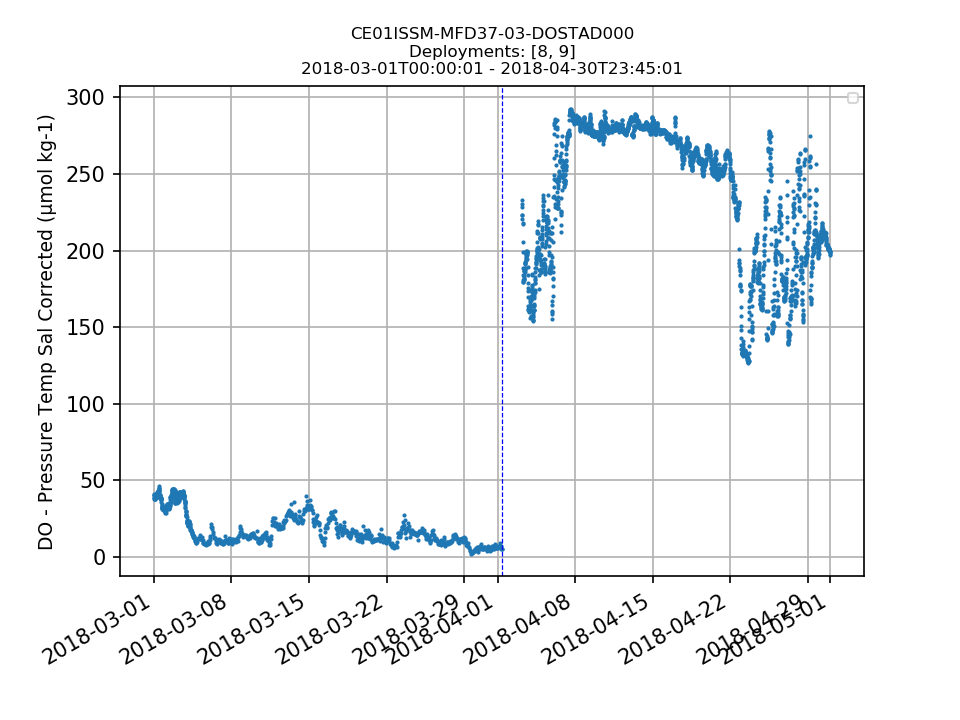
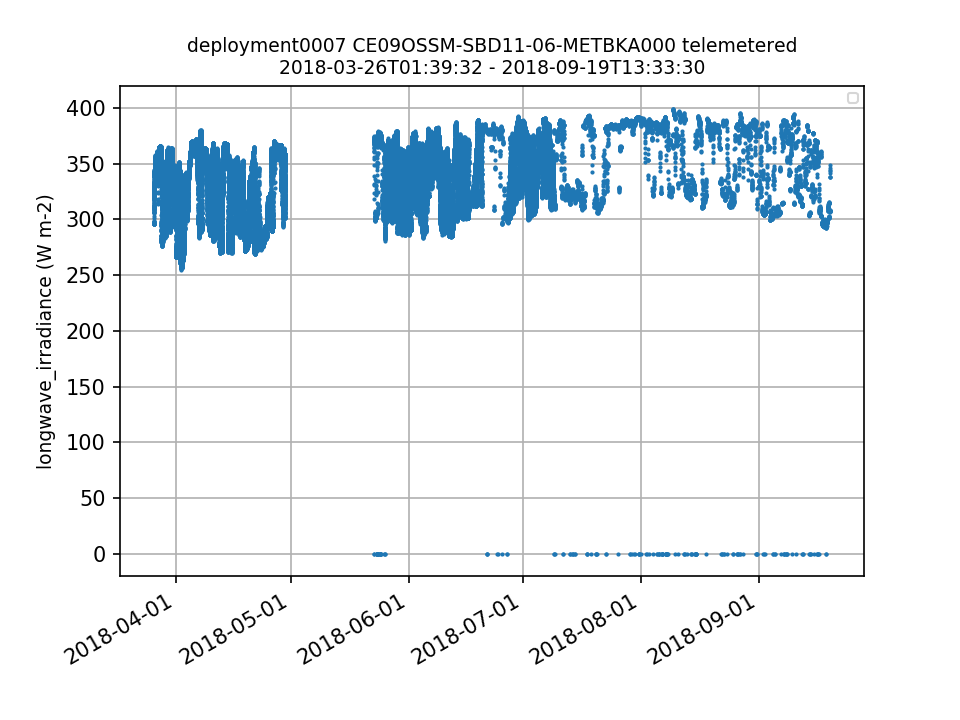
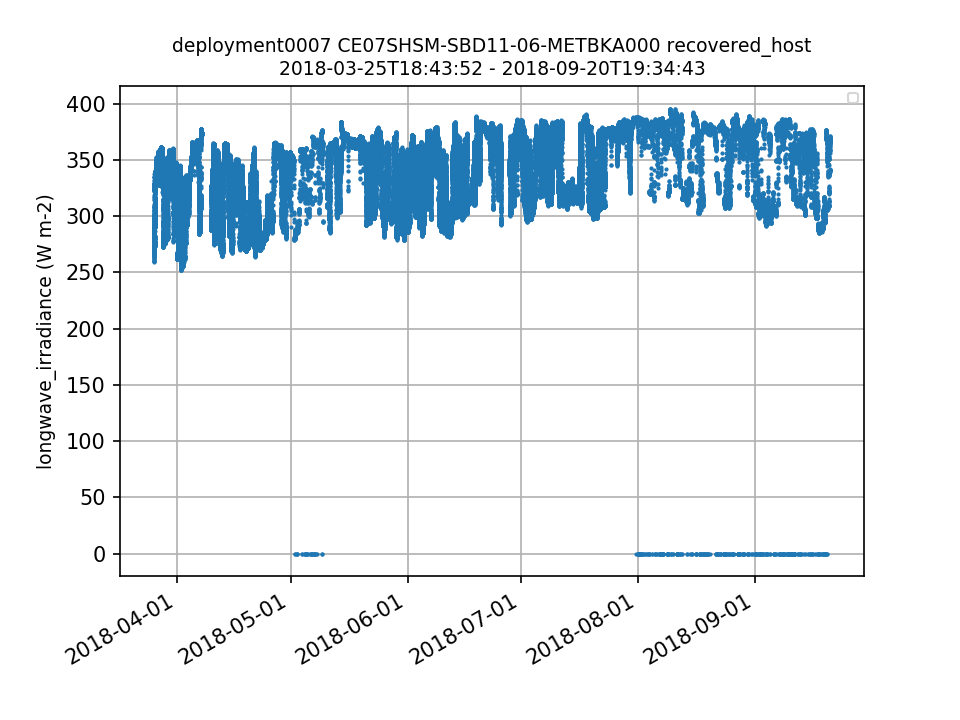
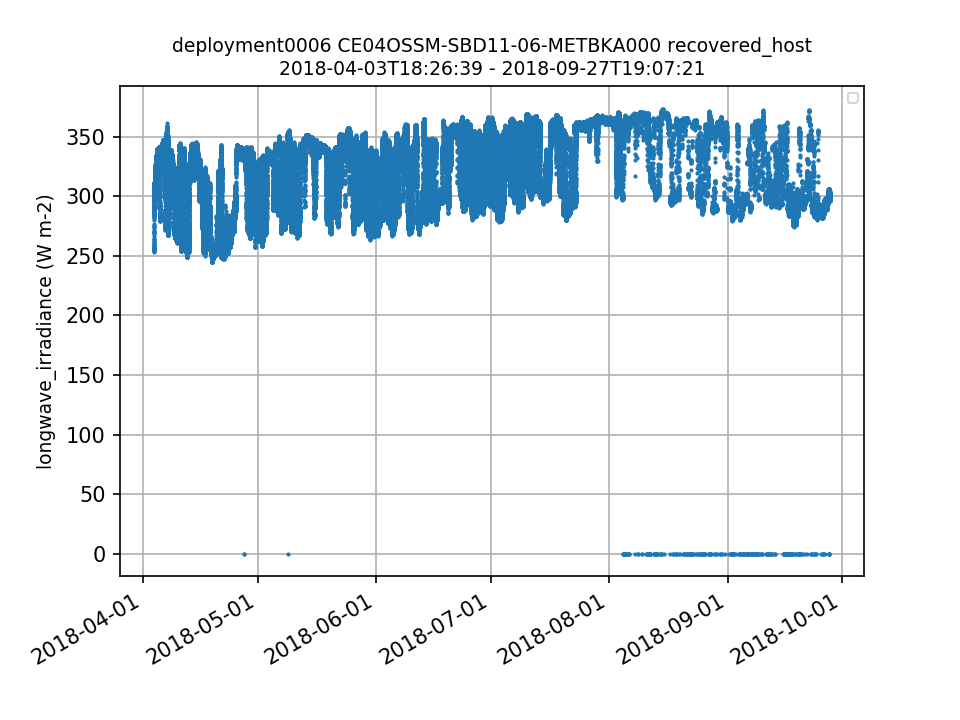
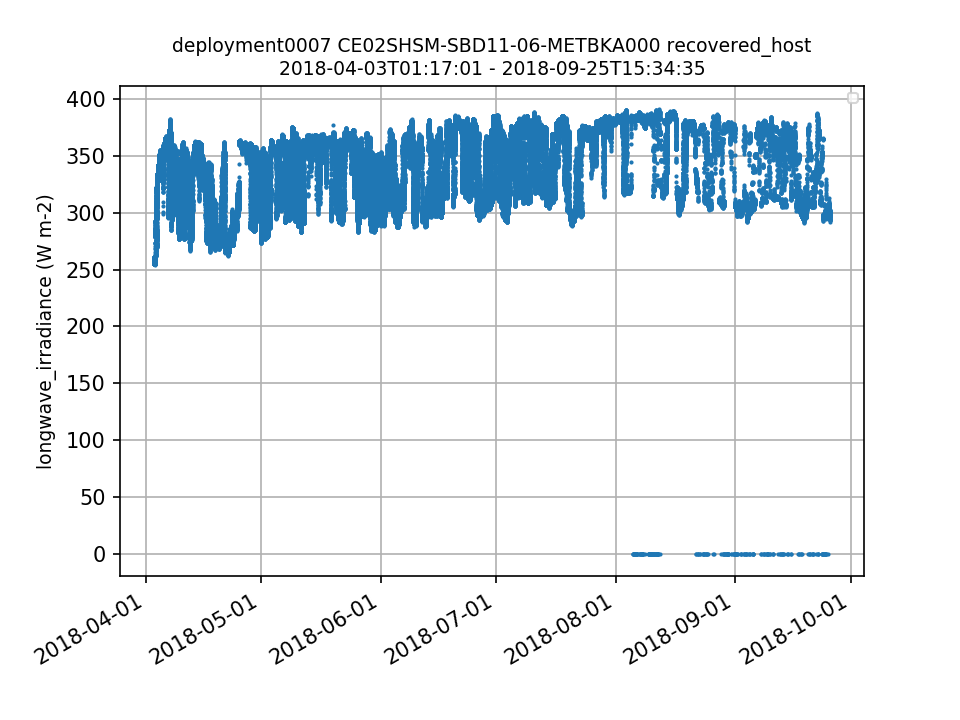


Figure 13. Dissolved Oxygen values at the end of deployment 8 are <50 umol kg-1 and jump to 150-200 umol kg-1 at the beginning of deployment 9 for CE01ISSM-MFD37-03-DOSTAD000.

*3.3 METBK*

Starting in the summer of 2018, all four Endurance METBKs exhibit 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.

b.

c.

d.

Figure 14. Suspect Endurance METBK data in the summer/fall of 2018 for (a) CE02SHSM, (b) CE04OSSM, (c) CE07SHSM, and (d) CE09OSSM.

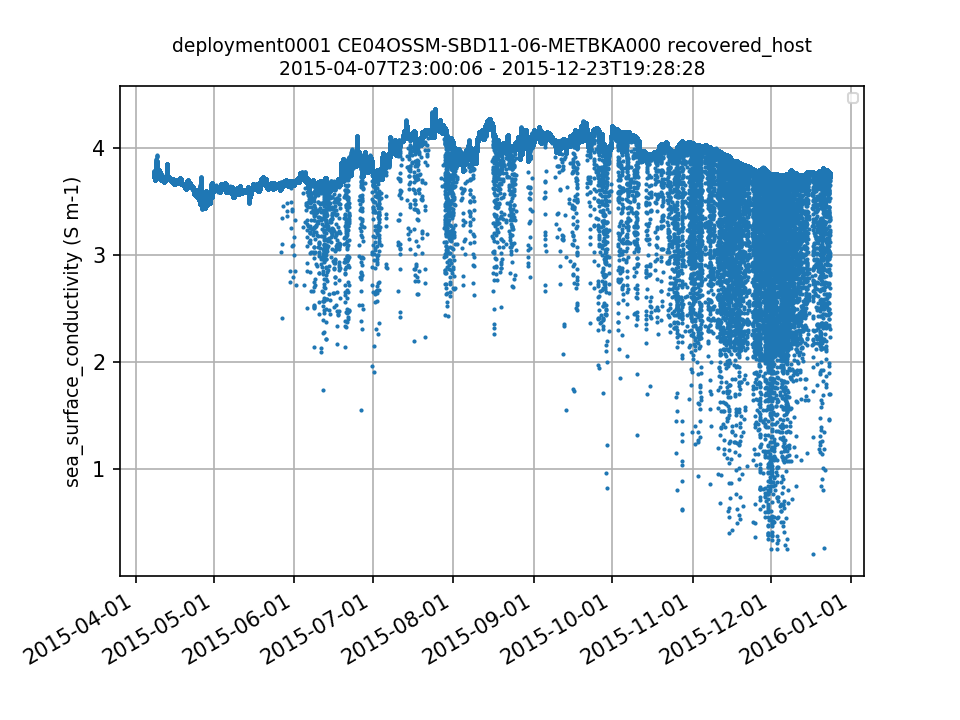


Figure 15. Issue with CE04OSSM-SBD11-06-METBKA000 deployment 1 Sea Surface Conductivity that requires an annotation.

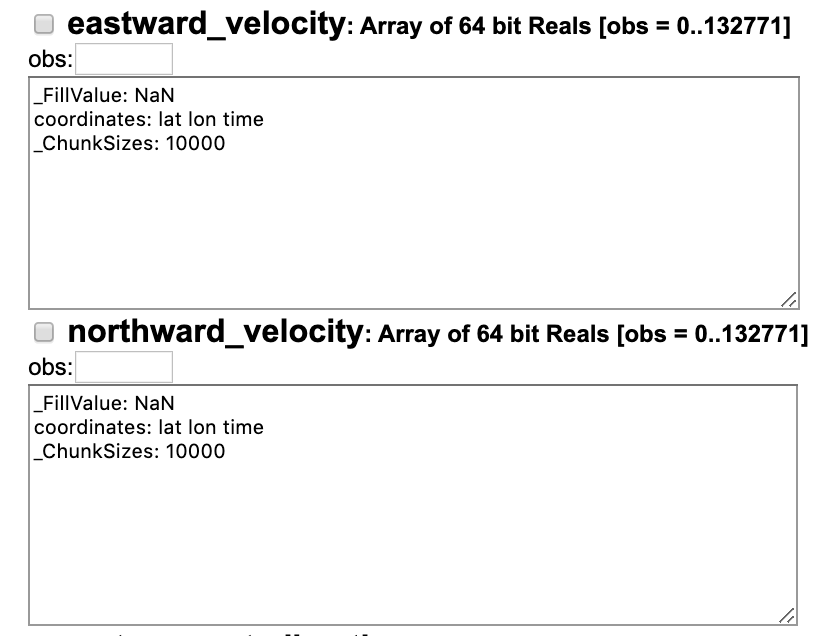


Figure 16. Screen shot of *eastward\_velocity* and *northward\_velocity* from the OPENDAP link for the deployment 4 CE07SHSM-SBD11-06-METBKA000 recovered\_host metbk\_a\_dcl\_instrument\_recovered showing the lack of any identifying metadata.

*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 the coordinate, or these data 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).

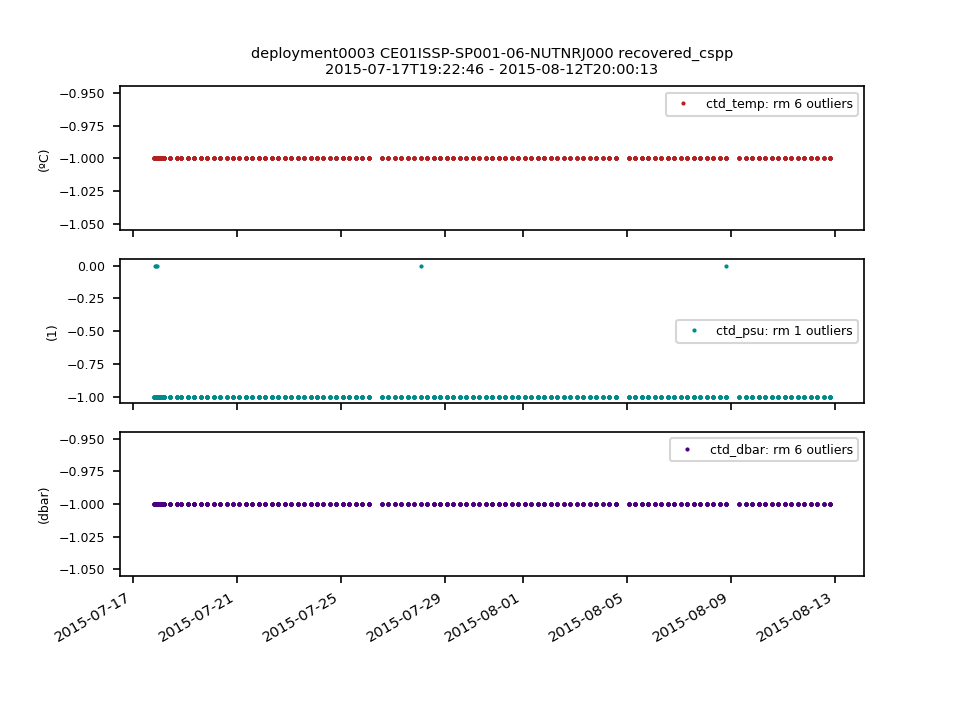


Figure 17. Incorrect values for *ctd\_temp*, *ctd\_psu*, and *ctd\_dbar* for deployment 3 of CE01ISSP-SP001-06-NUTNRJ000.

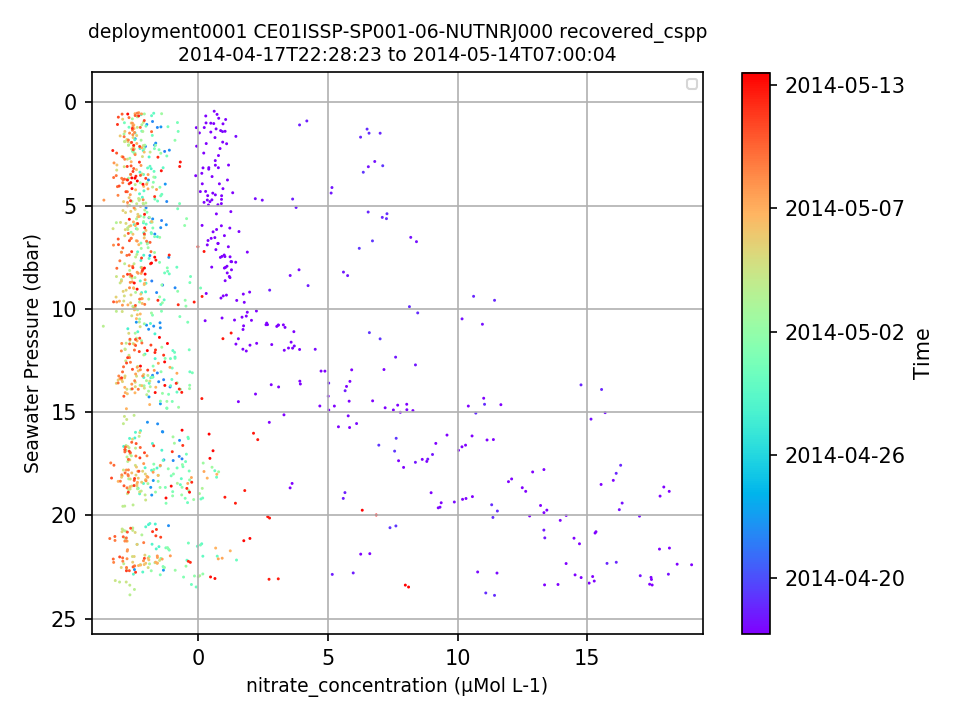


Figure 18. 70% of *nitrate\_concentration* values are negative for deployment 1 of CE01ISSP-SP001-06-NUTNRJ000.

**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 12. 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 |

*4.2 PCO2A*

**Redmine 14527:** The variable *pco2\_co2flux* is an array of fill values for every deployment of the Pioneer and Global PCO2A instruments (Table 13). According to the provenance.json files, there is an error in the pco2\_co2flux function in ion\_functions.data.co2\_functions.

Table 13. Pioneer and Global PCO2A instruments that don’t 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*

**Redmine 14504**: 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. 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*

**Redmine 13757**: 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, and was resolved in Dec 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 and the METBK calibration coefficients have been added to the system.

*5.3 Fluorometers on profilers: consistent data quality issues*

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.

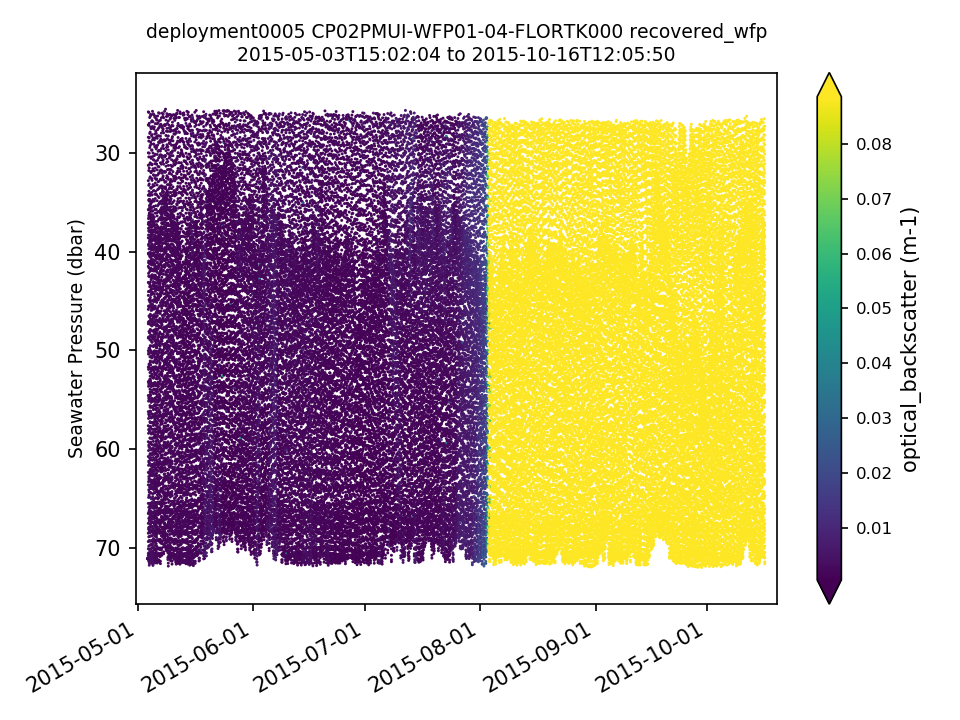


Figure 19. Example of data quality issue for fluorometers on Pioneer profilers.

**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 aren’t available. Also, annotation IDs 378-379 read “recovery options pending” – these annotations need to be updated.

**Redmine 13628**: GP03FLMB recovered\_host CTDMO data (12 instruments) are not available for download for deployment 5.

*6.3 Global Hybrid Profiler data issues*

**Redmine 13743:** 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 (either up or down) the offset between the values recorded for the same timestamp become increasingly more offset. When the profiler reaches the end of the profile, either at its shallowest or deepest 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.

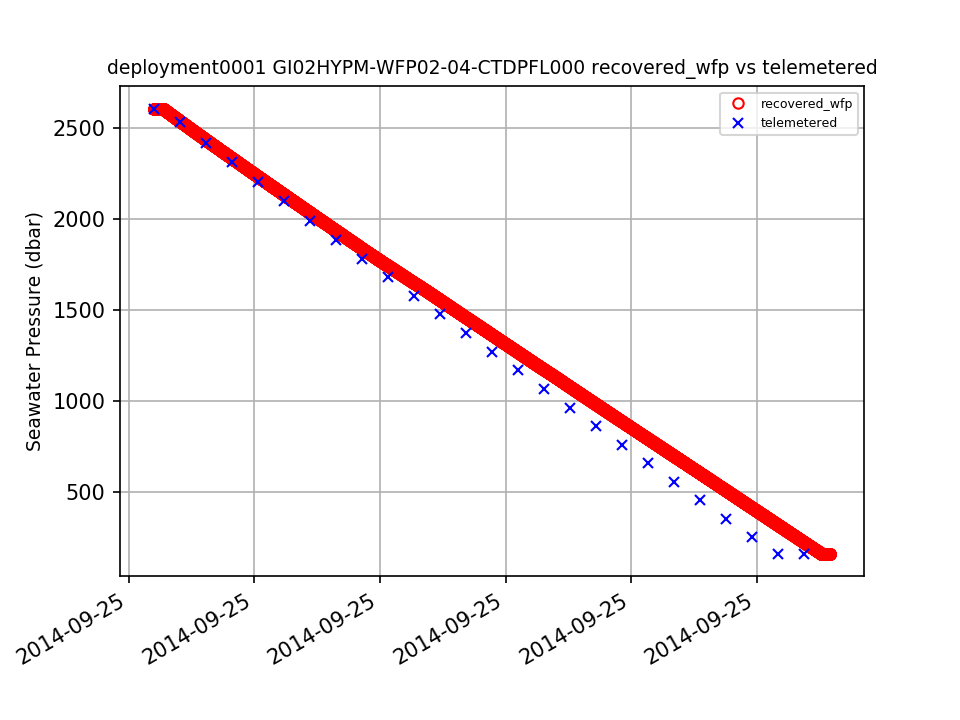
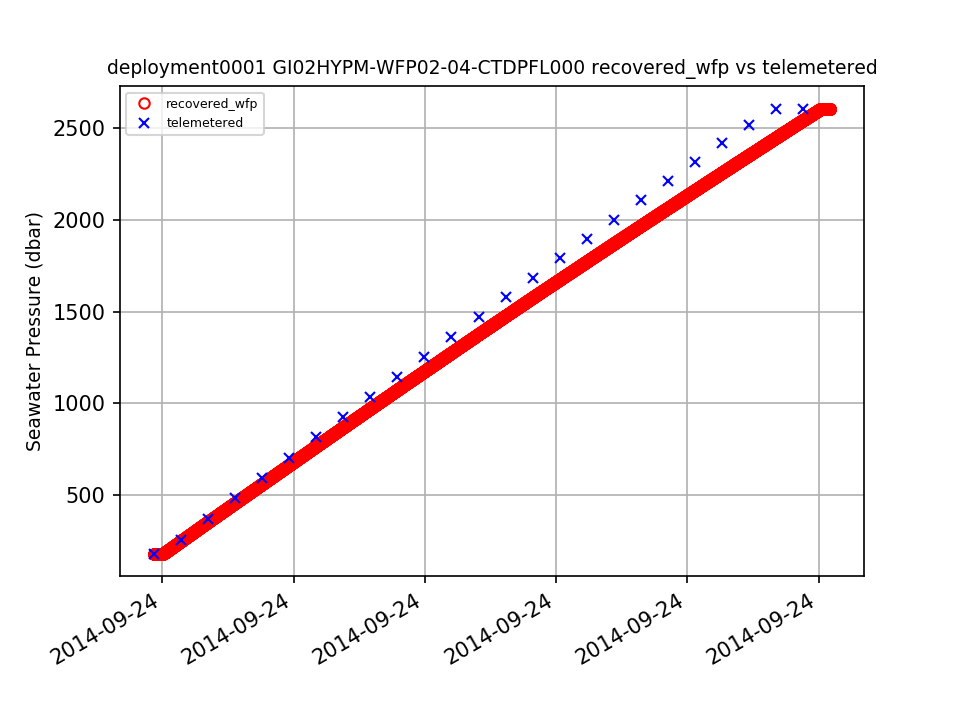


Figure 20. Recovered\_wfp compared to telemetered data for one down- and up-cast showing the increasing offset in pressure over the length of the cast.

**Redmine 13505:** 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.

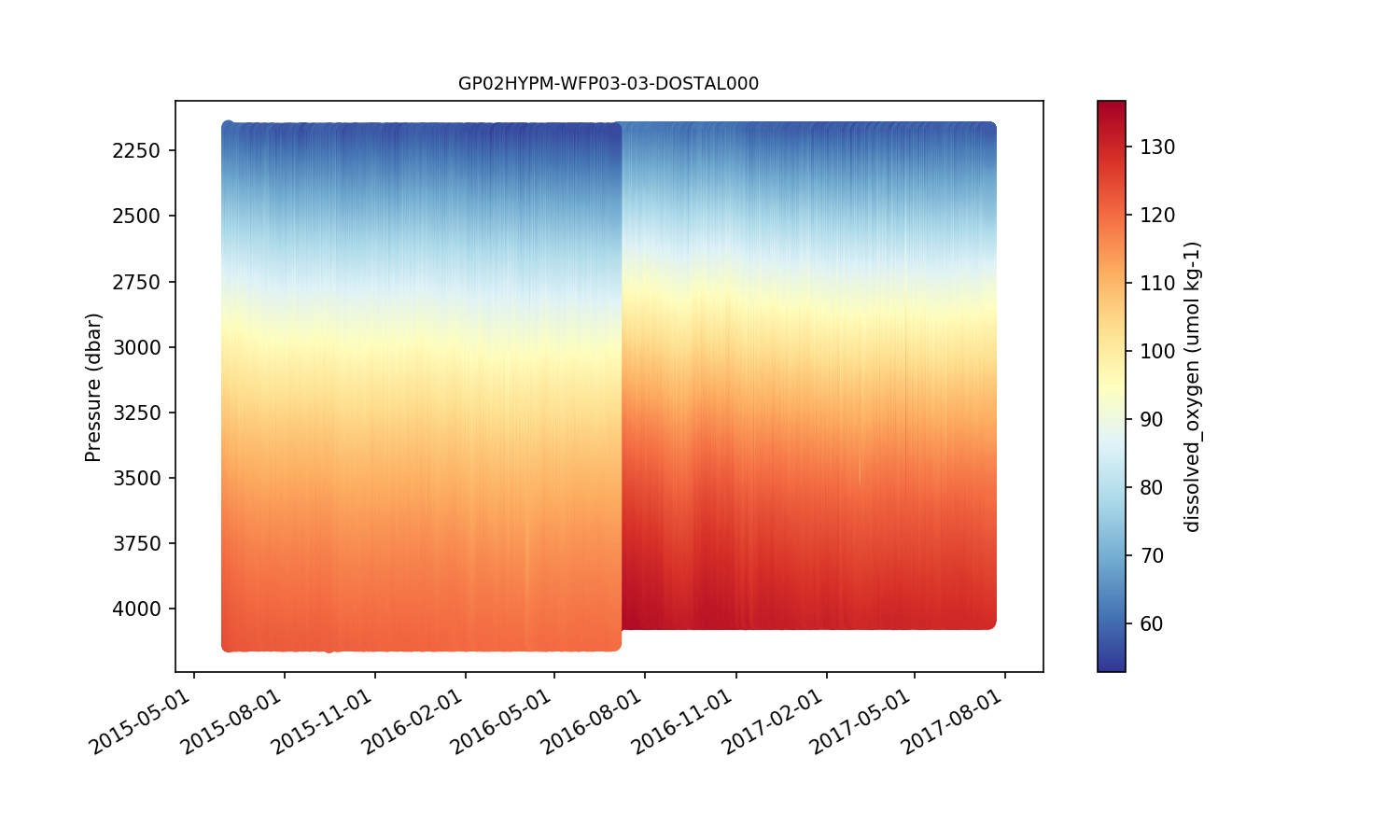
**

Figure 21. Unreasonable shift in Dissolved Oxygen between deployments 3 and 4 for GP02HYPM-WFP03-03-DOSTAL000

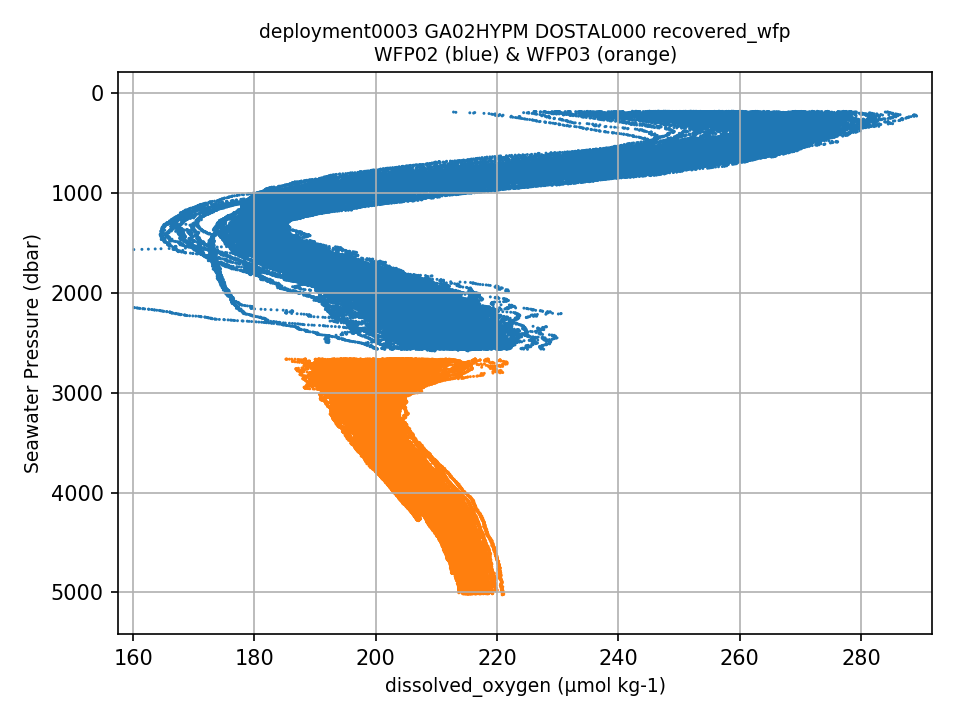


Figure 22. Example of the offset between values recorded by the Dissolved Oxygen sensors on the shallow (WFP02) and deep (WFP03) profilers.

*6.4 Values don’t match between delivery methods*

**Redmine 12573:** CTD data from recovered and telemetered delivery methods from the same instrument don’t match for several global CTD deployments (Table 13, Fig. 23). This could be related to calibration issues.

Table 13. Instrument deployments where data between different delivery methods don’t 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 |

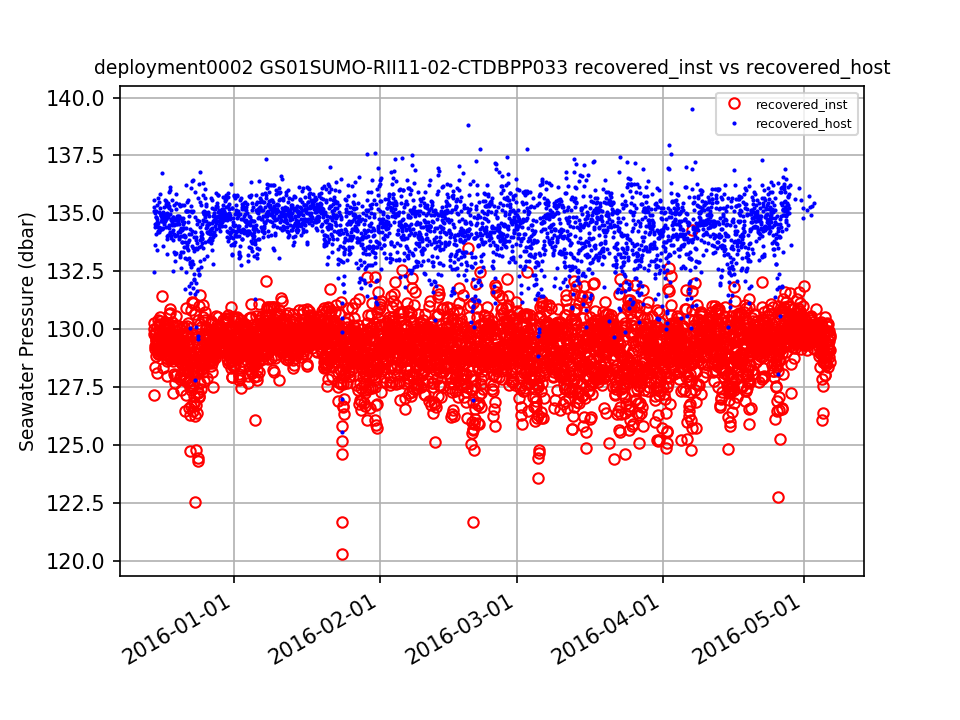


Figure 23. Example data from two different delivery methods (recovered\_inst and recovered\_host) from one instrument where the values are different for the same deployment.

**7. Cabled Array Issues**

*7.1 Profilers*

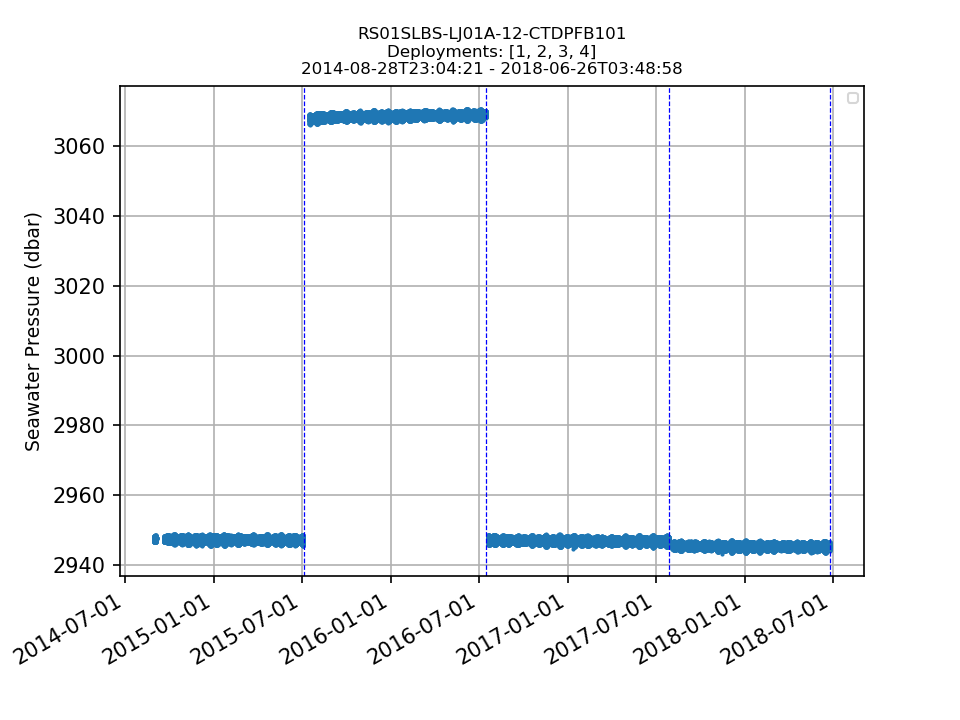
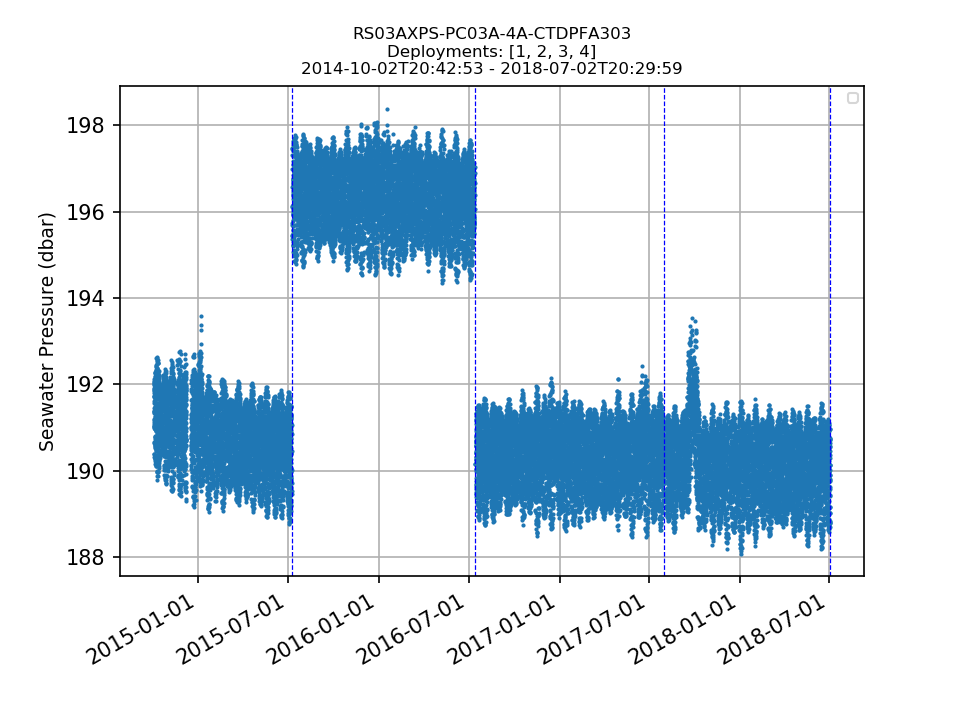
The Cabled Deep and Shallow Profilers are missing deployment depths in Asset Management. In addition, there have been several operational challenges regarding the Deep Profilers and the datasets should be annotated to alert users of the issues with these platforms.

*7.2 CTD calibration offsets*

Due to calibration errors for deployment 2 of several CTDs (Table 14), 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 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’re being applied properly.

Table 14. 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 ~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.

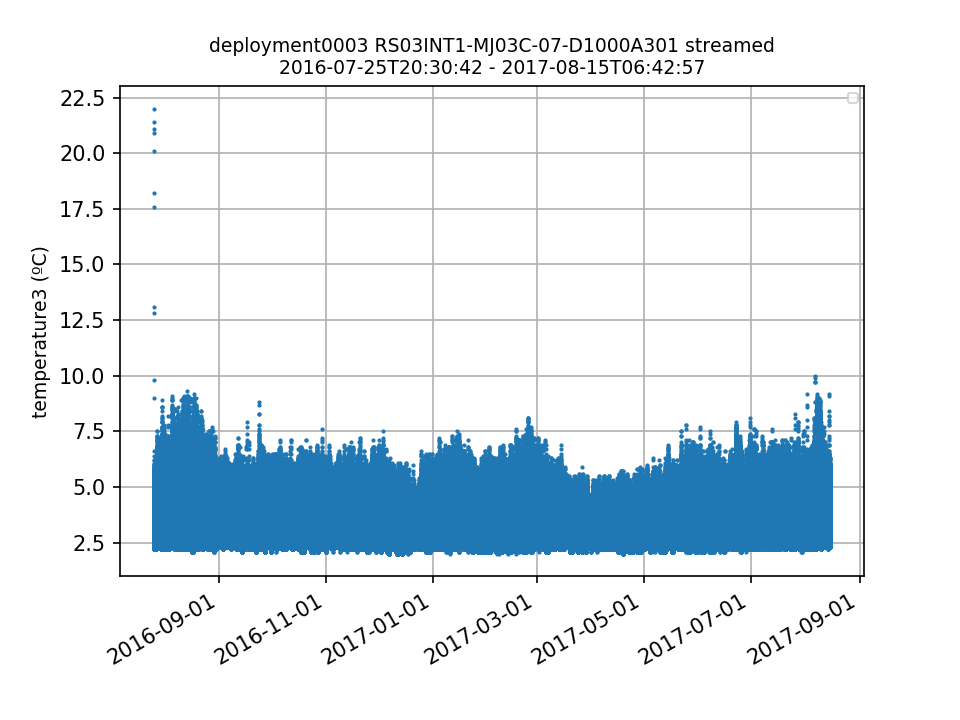


Figure 25. Values recorded by one of the thermistors on deployment 3 of RS03INT1-MJ03C-07-D1000A301 are not consistent with vent fluid temperatures.

*7.5 DOSTA – this section needs help, I can’t figure out how to explain it in writing*

Dissolved Oxygen data for the OOI 1.0 deployments are currently only available in the collocated CTD data stream (see Redmine #8662 for details). The data stream containing dissolved oxygen has been separated from the CTD data stream and OOI 2.0 dissolved oxygen data are now available in the DOSTA reference designator data streams. The OOI 1.0 dissolved oxygen data should also be ingested in DOSTA reference designator data streams so users don’t have to download two separate reference designators (CTD and DOSTA) to get different time ranges of dissolved oxygen data.

*7.6 HPIES*

**Redmine 14559**: the OOI is running an outdated version of THREDDS, and the NetCDF files for the echo\_sounding HPIES data stream can’t be accessed via the OPENDAP link in the OOI THREDDs server. The THREDDs server needs to be upgraded to have the capability to serve these data via the OPENDAP link. Note: the NetCDF files for this data stream can be downloaded and accessed locally.

*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*

**Redmine 14486**: The glider latitude and longitude variables were changed in Sept 2019 (see Redmine #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. All glider data were purged and re-ingested in Dec 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.

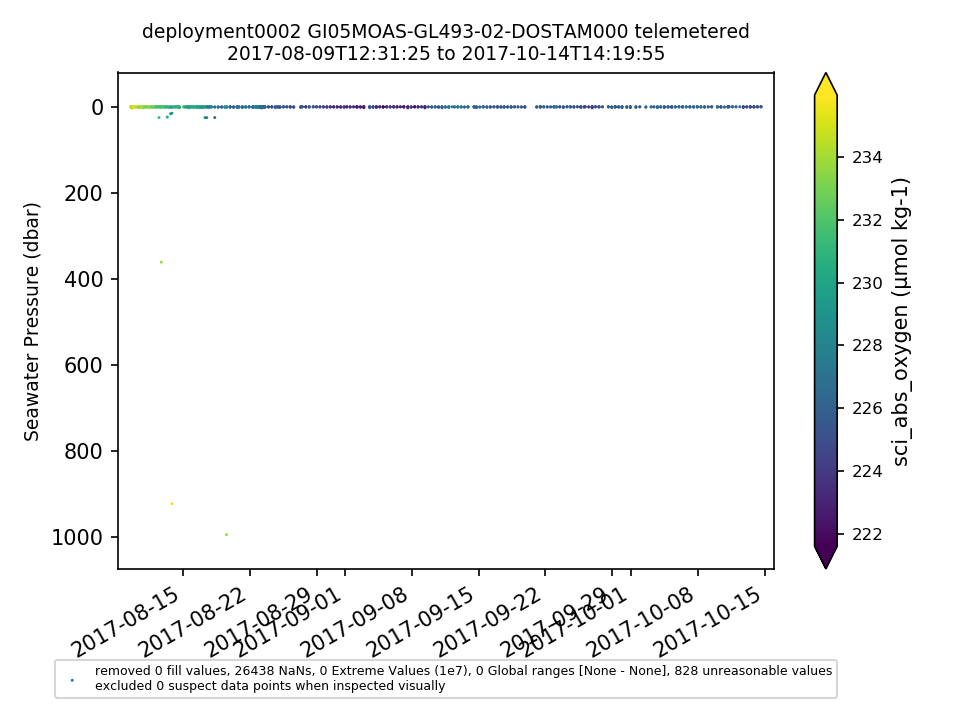
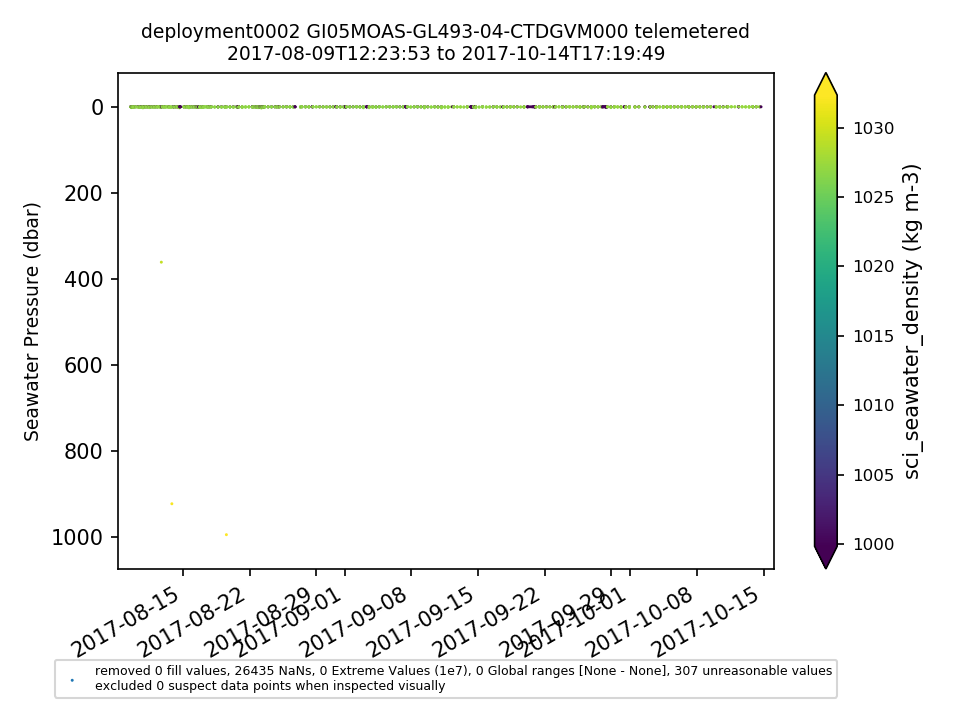
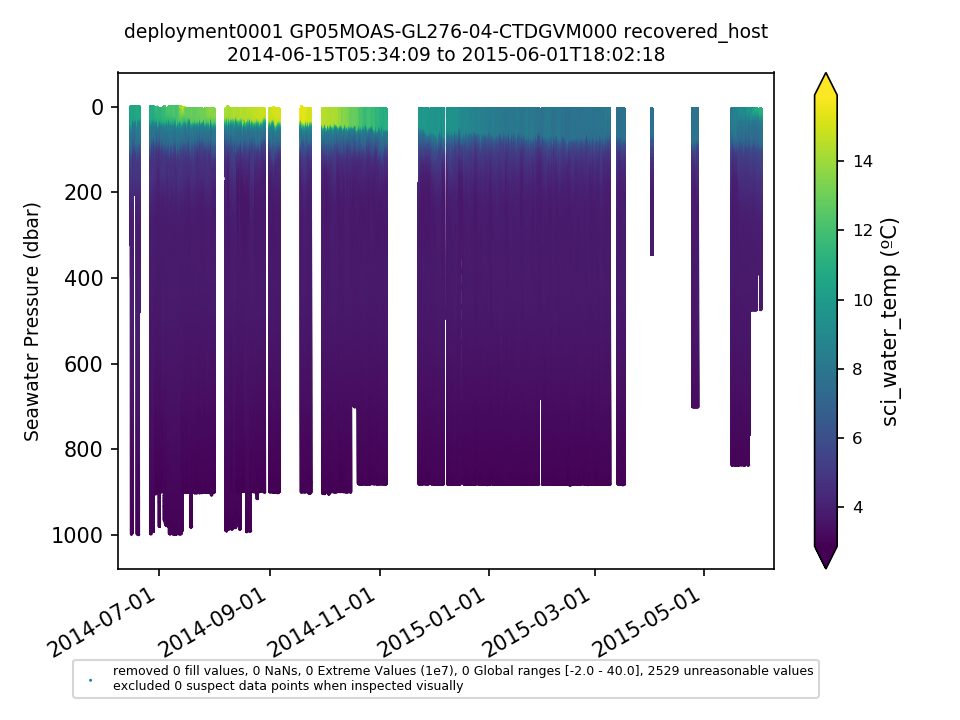
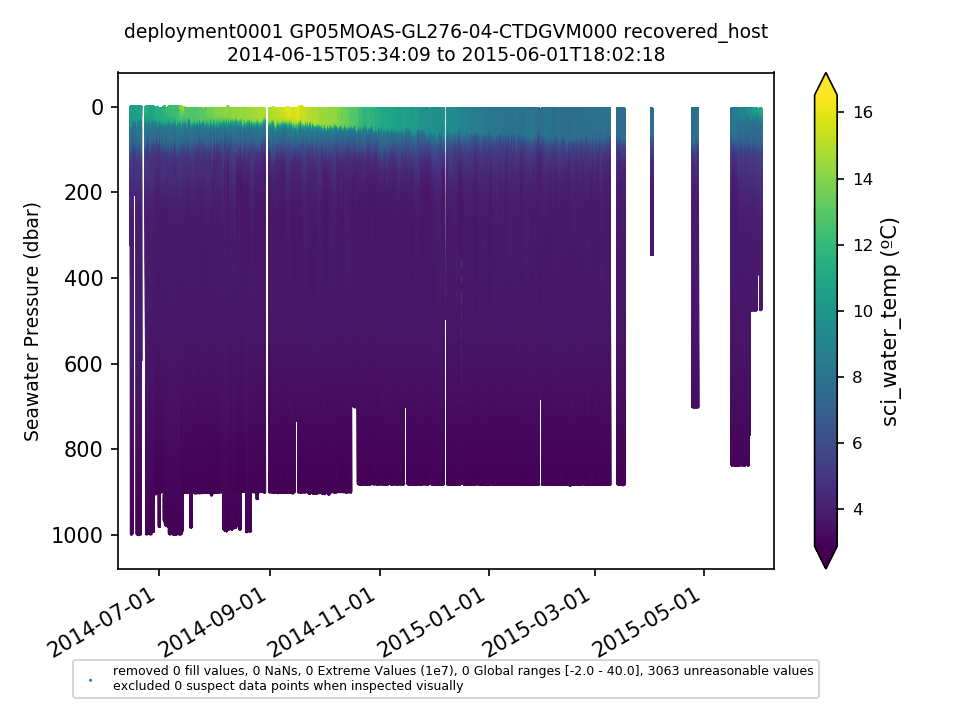


Figure 26. 98% of Seawater Density and Dissolved Oxygen values were NaNs for deployment 2 of GI05MOAS-GL493 after the re-assignment of the glider latitude and longitude variables.

**Redmine 14626**: after 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 just prior to NetCDF generation, and was fixed in a software release on March 3, 2020.

**Redmine 14654**: after the software release that fixed the recovered glider NetCDF file generation, a substantial amount of recovered data was 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 latitude and longitude are still missing from some glider NetCDF files (**Redmine 14657**).



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 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).

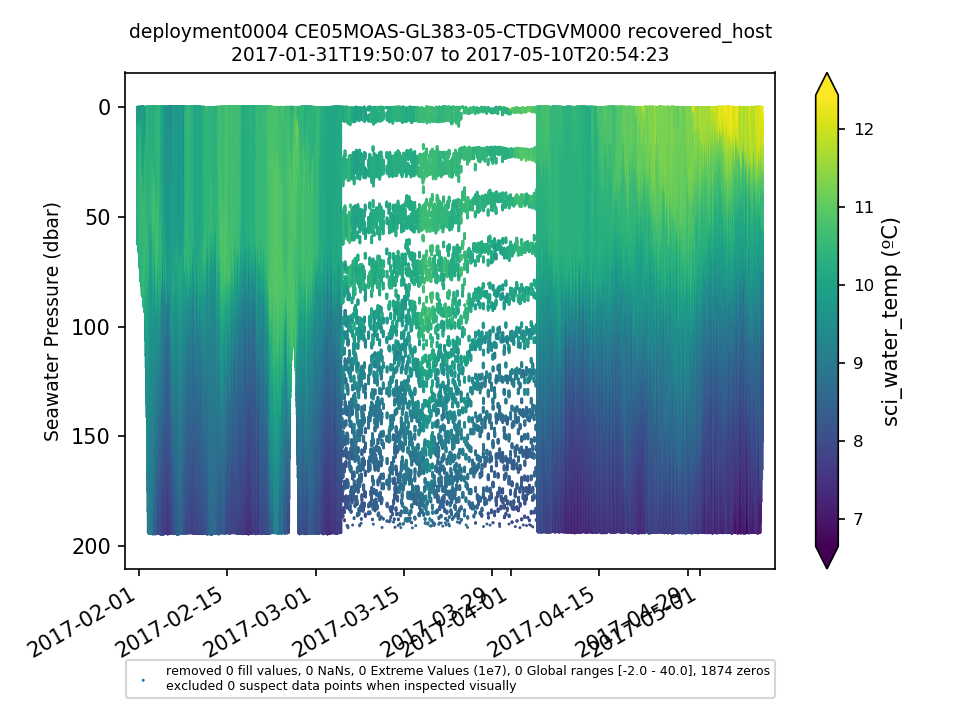


Figure 28. Apparent sampling pattern change in the middle of deployment 4 of CE05MOAS-GL383.

*8.3 PARAD*

For many glider deployments, the majority of PAR values are negative, and fail the Global Range QC test (Fig. 28). 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 15).

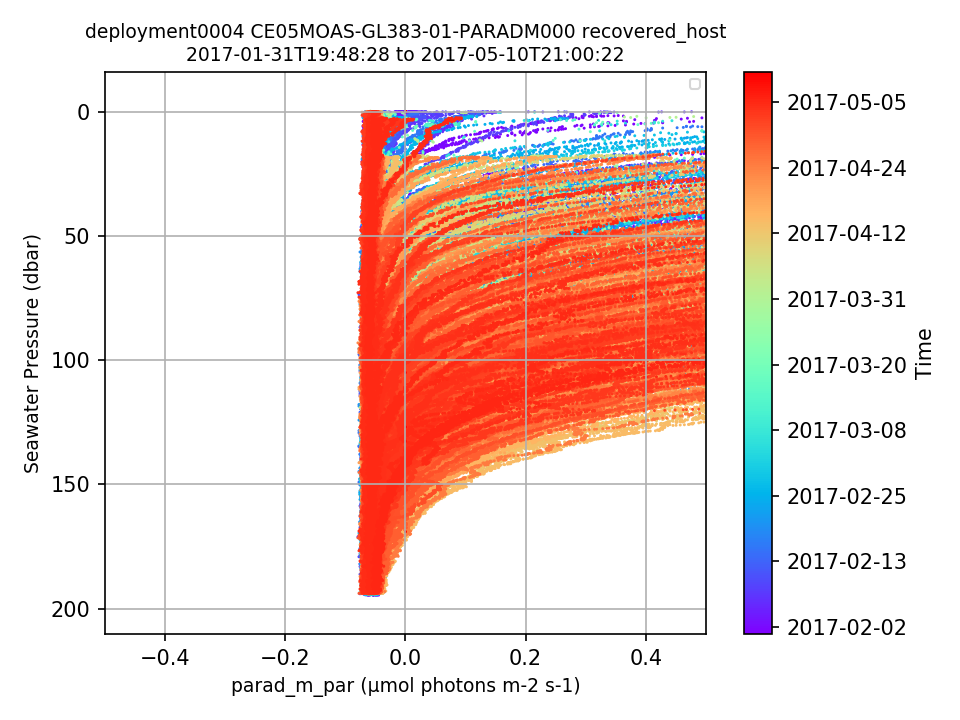


Figure 28. An example showing the majority (70%) of PAR data for one deployment are outside of global ranges (negative).

Table 15. 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 15), and CP05MOAS-GL336-02-FLORTM000 deployment 3 is missing CC\_scattering\_angle, CC\_measurement\_wavelength, CC\_depolarization\_ratio and CC\_angular\_resolution.

Table 15. 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. 29). 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.

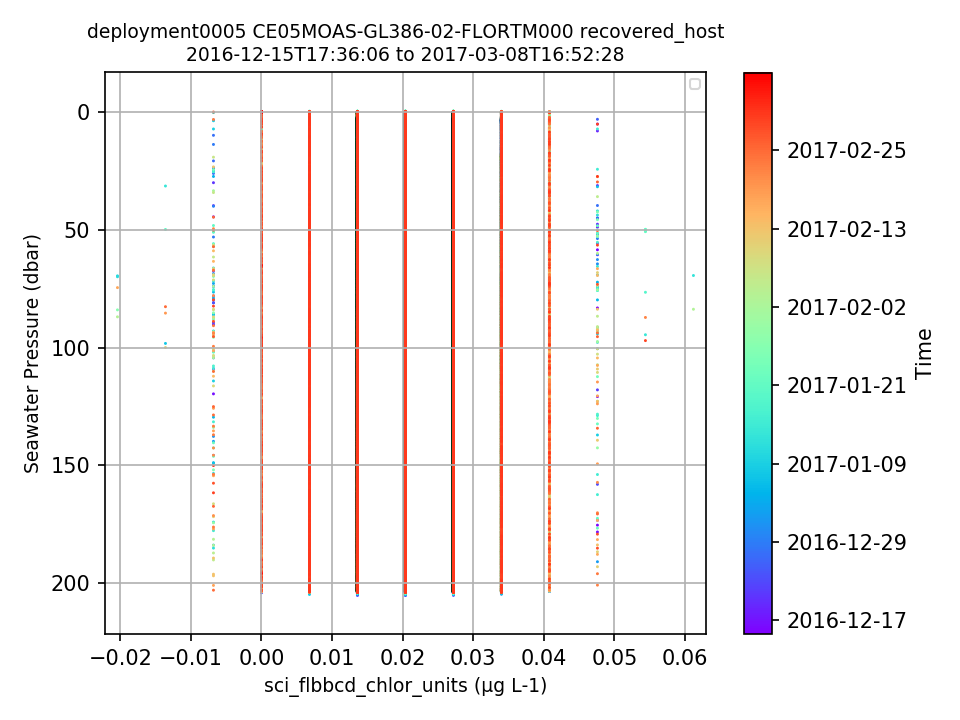
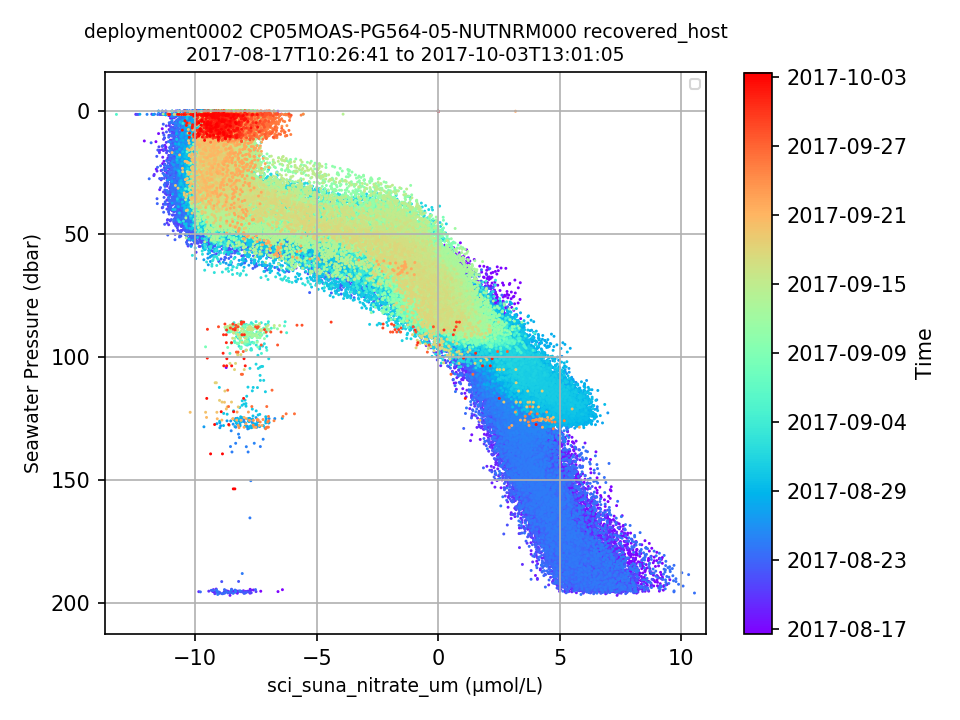
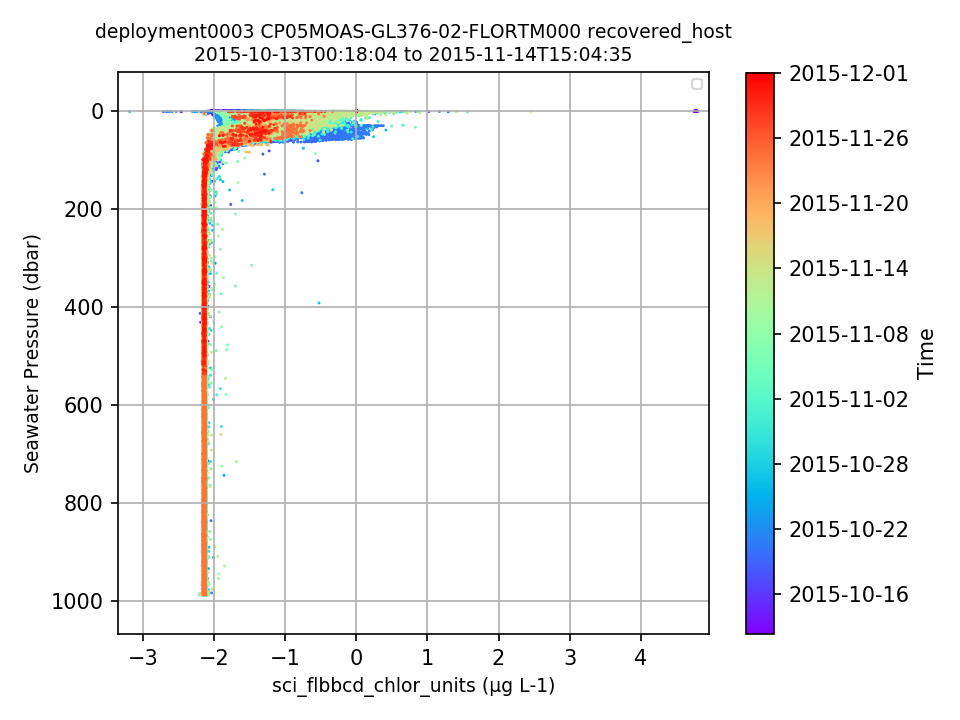


Figure 29. Chlorophyll-a data are close to zero for the entire deployment due to black electrical tape covering the sensor.

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. 30). These issues should be investigated and annotated.



a.

b.

Figure 30. 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.

**Table S1**. Endurance data that need to be ingested or annotated to explain why they aren’t available for download.

| Reference Designator | Delivery Method | Deployment(s) | 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 |  |
| 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 |  |
| 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 |  |

**Table S2**. Pioneer data that need to be ingested or annotated to explain why they aren’t available for download.

|  |  |  |  |
| --- | --- | --- | --- |
| Reference Designator | Delivery Method | Deployment(s) | 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 |  |
| 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 |  |

**Table S3**. Global data that need to be ingested or annotated to explain why they aren’t available for download.

|  |  |  |  |
| --- | --- | --- | --- |
| Reference Designator | Delivery Method | Deployment(s) | 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 |  |
| 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 S4**. Cabled data that need to be ingested or annotated to explain why they aren’t available for download.

|  |  |
| --- | --- |
| Reference Designator | Deployment(s) |
| 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 S5**. 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(s) |
| 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 S6**. 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(s) |
| 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 S7**. 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 |