

# PERSPECTIVES FOR ACCURACY AND QUALITY ASSURANCE OF CTD & BIOGEOCHEMICAL DATA STREAMS FROM AUTONOMOUS PLATFORMS

Ian D. Walsh<sup>(1)</sup>, Jochen Klinke<sup>(2)</sup>, Kim Martini<sup>(3)</sup>, David Murphy<sup>(4)</sup>

<sup>(1)</sup> *Sea-Bird Scientific, 620 Applegate St, Philomath, OR 97370, USA, Email: iwalsh@seabird.com*

<sup>(2)</sup> *Sea-Bird Scientific, 13431 NE 20<sup>th</sup> St, Bellevue, WA 98005, USA, Email: jklinke@seabird.com*

<sup>(3)</sup> *Sea-Bird Scientific, 13431 NE 20<sup>th</sup> St, Bellevue, WA 98005, USA, Email: kmartini@seabird.com*

<sup>(4)</sup> *Sea-Bird Scientific, 13431 NE 20<sup>th</sup> St, Bellevue, WA 98005, USA, Email: dmurphy@seabird.com*

## ABSTRACT

The analysis of near-real time data from autonomous platforms poses a particular challenge in that there is a near-constant stream of incoming new data. In order to prevent data overflow, an effective quality assurance mechanism must become part of the data delivery. As future Autonomous Lagrangian Profilers and Sensors (ALPS) quality assurance (QA) systems move closer towards real time, all factors of uncertainty within the system must be measured so that efforts are made to constrain the total uncertainty. Here we describe elements of an automated QA system that must be considered as we scale up the data collection into the biogeochemical realm and extend the measurement to the deep ocean.

## 1. INTRODUCTION

ALPS are at the core of the Argo. This program features over 3800 profiling floats that roam the upper 2000 m of the ocean to measure its salinity and temperature. However, about half of the ocean's volume is contained in depths greater than 2000 m. Systematic observations deeper than 2000 m rely solely on ship-based measurements and deep moorings with sensor arrays. To date, less than 5% of T/S profiles in the World Ocean Database are deeper than 4000 m. In addition, these data are biased toward the Northern Hemisphere.

The resulting sparse spatial and temporal coverage is not sufficient to describe ocean circulation in many regions or decadal variations. In order to close the gaps of global heat budgets and freshwater storage as well as improve the knowledge of steric sea level changes, sampling of the deeper ocean is necessary. Efforts to extend the observations to deeper depths has led to the inception of the Deep Argo program. Over recent years, four models of Deep Argo floats have been developed and used in pilot deployments (Zilbermann and Maze, 2015). These instruments can profile to up to 6000 m and will eventually provide data quality that rivals even that of WOCE/GO-SHIP observations.

Fleets of autonomous vehicles deliver data streams that quickly dominate even the largest traditional data sets. Therefore, an effective quality-assurance mechanism must become part of the data delivery system, or else there will always be a growing data overrun. A timely assessment of data quality is dependent on an understanding of the nature of the sensors themselves. This includes quantification of the change in the sensor response and signal over the expected life of the instrument. Further, the sensor response must be quantified with respect to the intrinsic scale of the measurement and the target. Balancing the scale of the sensor and the target is particularly important for biogeochemical parameters and biological organisms, as the targets are found intermittently and in patches.

The Argo program has built a system that works for the basic physical parameters that are found continuously and globally. Describing the functionality of the sensors used on autonomous vehicles, particularly with respect to the expected performance criteria over the expected lifetime of the sensors, is critical in ongoing analysis of anomalies in data streams to allow partitioning of data between changes in the sensor versus changes in the environment.

Here we describe elements of an automated quality-assurance system, which must be considered as an effective future ALPS QA system must move closer towards real time, and all elements of uncertainty within the system must be measured so that efforts are made to constrain the total uncertainty. First, we focus on recent advances in the core physical parameters measured by the Argo program, including characterization of the SBE 41CP CTD dynamic response, and assess the performance of the SBE 61 CTD instrument during Deep Argo pilot deployments. Then, we consider the requirements for sampling and QA as we scale up the data collection into the biogeochemical realm. Finally, we discuss the implications of these developments for observations within the framework of the Argo program.

## 2. CTD DATA STREAMS

### 2.1. Core Argo CTD accuracy

The SBE 41CP pumped CTD has been the standard instrument used to measure the salinity and temperature on traditional Argo floats. The majority of the transmitted profile data achieve temperature and salinity accuracies of  $\pm 0.002$  °C and  $\pm 0.01$  psu respectively, with a depth uncertainty of  $\pm 2.4$  m. With the application of delayed mode corrections, e.g., to remove pressure drifts, about 90% of the data fall within these error bounds. One factor that currently limits the accuracy of Argo data is the averaging of the uncorrected CTD data into 1-second bins before transmission via satellite. In principal, applying corrections to the CTD data before binning could yield less noisy and higher resolution data.

Through a series of laboratory experiments performed in the Stratified Tank at Woods Hole Oceanographic Institute (Schmitt et al., 2005), Sea-Bird has made great strides in characterizing the dynamic response of the temperature and conductivity sensors. By profiling through a nearly step-change in temperature and salinity, corrections for thermistor thermal mass, the time lag between temperature and conductivity, and conductivity cell thermal mass are determined for Argo CTDs. Applying the corrections reduces errors caused by the response time of the instruments, such as salinity spiking and smearing when profiling through large changes in temperature and salinity (Figure 1).

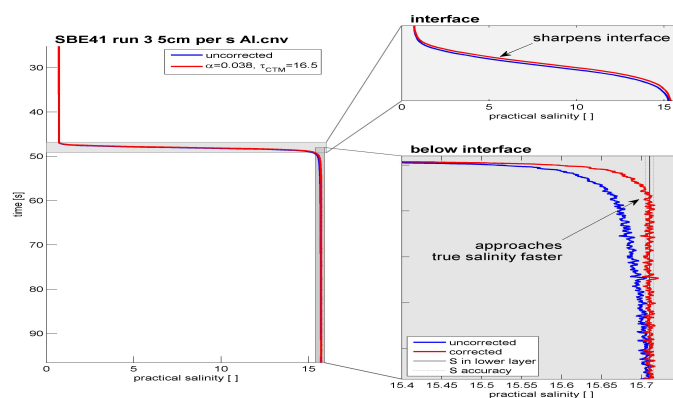


Figure 1. SBE 41CP salinity profile of a 15 psu step-change in a stratified tank. Uncorrected data (blue) show a slower response than data that are corrected for thermal mass of thermistor and conductivity cell (red).

Such improved response times allow the resolution of finer structures in the water column, especially in areas where the temperature and salinity are changing rapidly. Further investigation and quantification of these cell thermal mass terms are needed before real-time correction algorithms on-board the floats become conceivable, as such data manipulations would, in essence, be irreversible. The potential benefit of these adjustments for the data quality will have to be weighed against the inherent risk of rendering the data useless should mistakes occur in processing. In light of these risks, an effective implementation of real time QA will be essential for future ALPS.

### 2.2. Deep Argo CTD accuracy

Deep SOLO floats profile from the surface down to 6000 m, deeper than traditional Argo floats and most shipboard systems. These floats are equipped with a new generation CTD sensor that is specifically designed to meet the ambitious accuracy goals of Deep Argo. Targeted are uncertainties of  $\pm 0.001$  °C for temperature and  $\pm 0.002$  psu for salinity, and a pressure accuracy of  $\pm 3$  dbar. As stated earlier, these targets actually exceed even those achieved by WOCE/GO-SHIP observations.

Since 2014, 14 Deep SOLO floats have been deployed by Scripps Institution of Oceanography in the southern Pacific east of New Zealand (Zilbermann and Maze, 2015) with SBE 61 CTDs (Figure 2). The sensors have shown to be remarkably stable throughout the deployment. Salinity drift computed along isotherms at depths greater than 5200 m, where water mass properties change very little, is less than 0.002 psu for all CTDs shown. Salinity bias at depths greater than 4000 m seen in initial deployments of SOLO floats 6002 and 6003 are corrected in deployments after 2016.

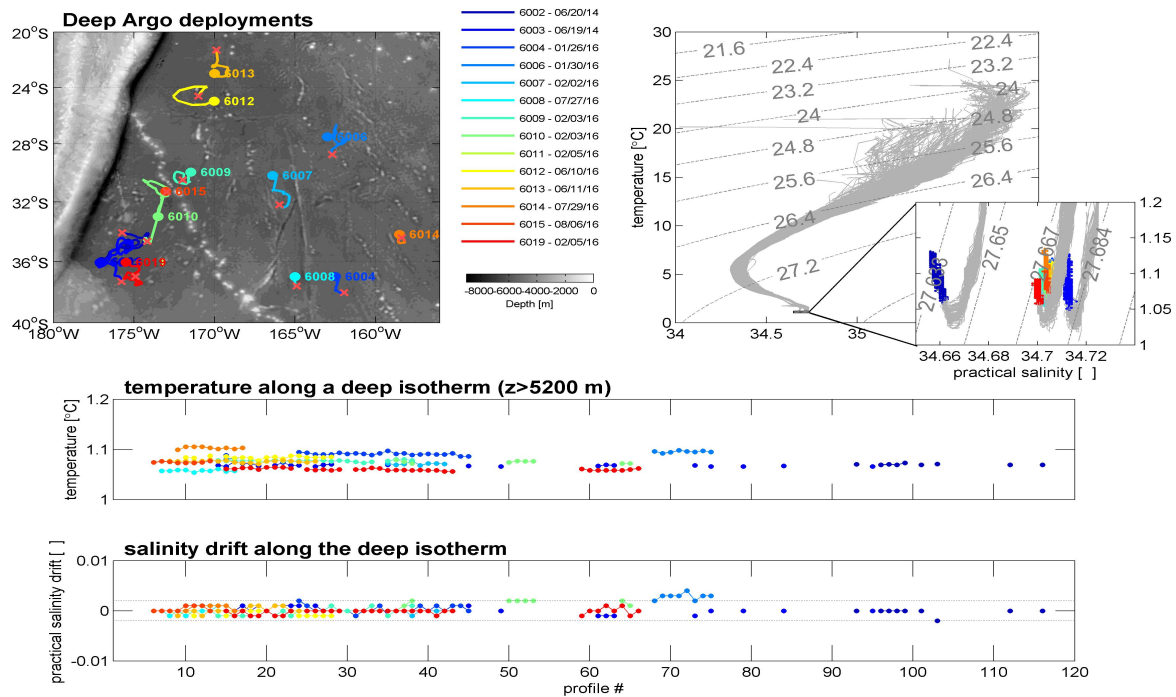


Figure 2. Temperature and salinity drift of Deep SOLO floats along deep isotherms.

Analysis of the pre- and post-cruise calibrations illustrates the excellent stability of the SBE 61 temperature sensor (Figure 3). Comparison of the SBE 61 profile data with the shipboard SBE 9plus CTD measurements shows that the achieved temperature accuracies are already within the specified range of Deep Argo accuracy goals. With regard to salinity, one Deep SOLO (6003) showed a small salty bias relative to shipboard SBE 9plus and historical CTD data, while the other Deep SOLO (6002) exhibited a larger fresh bias with depth-dependency. After removing initial biases, average salinity uncertainties ( $\pm 0.004$  psu) and pressure errors ( $\pm 4.5$  dbar) are approaching the desired target accuracy. However, continued efforts are needed to obtain a full characterization of the SBE 61 CTD and understand biases and depth dependency observed in the salinity data during the pilot deployments.

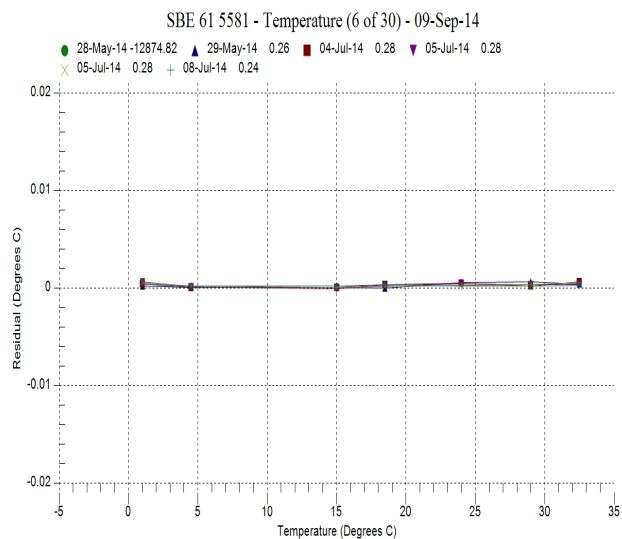


Figure 3. Pre- and post-cruise temperature calibrations for Deep SOLO SBE 61 CTD S/N 5581.

### 3. BIOGEOCHEMICAL DATA STREAMS

The flexibility of the current generation of float sensor packages provides an opportunity to craft mission-specific sampling schemes that balance the collection of data for specific sampling goals with the practicalities of float operation. However, as the number of sensors increases and the breadth of mission parameters expands, it becomes more difficult to optimize data sequencing and reporting.

Autonomous floats operate within constraints of battery life and data transfer rates. For simplicity of data transfer and handling, most float data sets are transmitted after binning on pressure. Within a given pressure bin, different instruments will be sampling within a particular defined sequence. Any particular sampling sequence

should be balanced toward minimizing energy consumption while maximizing the transmitted data accuracy of each instrument.

For buoyancy-modulated floats, there is an important feedback between ascent rate and data density. Data from each instrument is recorded as the float rises according to the programmed instrument sequence (Figure 4). The ascent rate of the float therefore determines the data density within the profile, as the instruments are driven by the time sequence. The ascent rate modulates the number of subsamples within each individual transmitted bin. The end result is that when gradients in pressure decrease the ascent rate of a float, there is an increased vertical sampling rate and local increases in the number of subsamples per pressure bin. Since many biogeochemical gradients coincide with density gradients, the net result is an increase in data density precisely where it is most

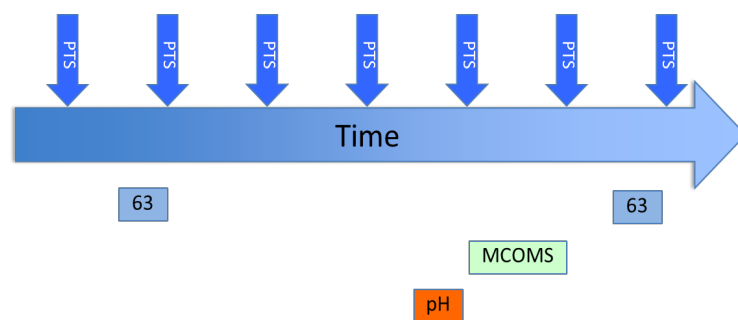


Figure 4. Diagram of data collection from a suite of sensors on a Navis profiling float. As the sensor load on autonomous vehicles increases, the optimization parameters for the suite of instruments becomes increasingly complex. Pressure, temperature, and conductivity (PTS in the diagram) are measured at 1 Hz, while serial instruments can have independent data rates up to 1 Hz. The Navis float controller can accept a range of sampling schemes, with each instrument having its own power and data parameters. This allows for instruments with longer equilibrium times to be continuously powered, while instruments with faster response times can be powered down between samples. In this example, the SBE 63 dissolved oxygen and pH sensors are continuously powered while the MCOMS optics instrument is put to sleep between each sample.

important to collect higher-density biogeochemical data.

While this is useful in and of itself, for particle dynamics and export studies it is important to have some information on the variance of the data observed (Walsh et al., 1992; Briggs et al., 2011; Poteau et al., 2016). With binned data, some of the variance is lost in the binning process. In most cases, autonomous float data sets consist only of the transmitted data, making analysis of the variance of the data difficult. As part of Sea-Bird Scientific's development work on optical and biogeochemical sensors for profiling floats, a float (Navis BGCi 0039) was deployed and recovered that was equipped with internal memory to capture the full set of sample data taken during the profiling mission. This data set allows us to compare the 'raw' data and the transmitted, and determine the impact of data sequencing and binning on the data.

Navis BGCi 0039 was equipped with an SBE 41N CTD measuring conductivity, temperature, and pressure (depth), an SBE 63 Optical Dissolved Oxygen sensor integrated within the CTD flow path, and an MCOMS optical instrument that provides chlorophyll a fluorescence and backscattering at 700 nm. The Navis float can be expanded with additional sensors; in this case an experimental pH sensor was incorporated into the float. The float was deployed off Hawaii in 2014 and recovered after five months and one hundred profiles. The transmitted data in the upper 1000 m of the water column was binned on board and transmitted at two-meter resolution. Data from the float's internal memory was processed upon recovery to the same two-meter bins, with the addition of the number of subsamples for each sensor added to each data line.

The transmitted and internally recorded data sets were compared visually by overlay plotting and by difference. For the physical parameters, i.e., temperature and salinity, there was no effective difference between the two data sets. For the chlorophyll fluorescence data, there was no effective difference between data sets except at the peak of the chlorophyll maximum, where the recorded data showed higher peak values.

The recorded backscattering data contained many more spike data points than the transmitted data (Figure 5, upper panel). Assuming the 'spikes' represent relatively rare large particles, the transmitted data undercounts the

large particles. Since the rare large particles are significant to the vertical flux of particles that in the open ocean are almost entirely of biological origin, systemic undercounting of large particles will result in underestimates of the vertical flux of carbon.

To reconcile the backscattering data sets, we applied a median filter to both the transmitted and recorded data sets, and then split the data into two fractions (small and large) by subtracting the median filtered data from the original data set (Figure 5, lower panel). As expected for this environment, the median filtered data were substantially the same for both the transmitted and recorded data sets. We then applied a sharpening algorithm that assumes that the difference between the raw and the median data was ascribable to a single data line by dividing the large particle value by the number of lines in the bin and adding that back to an individual line. This methodology was highly effective in generating from the transmitted data a resultant data set that had a similar number of spikes as well as a similar range of absolute values as the recorded data set.

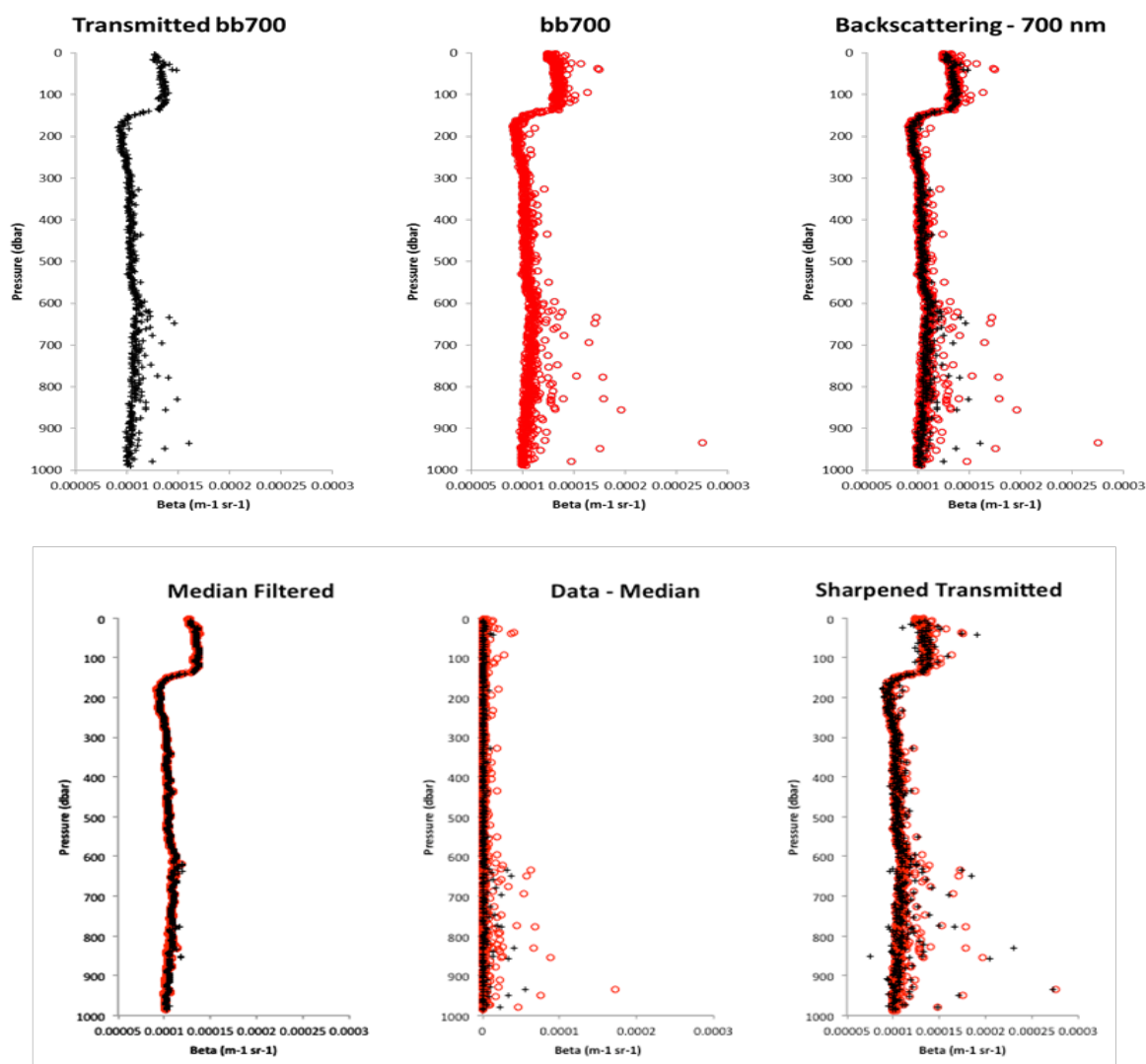


Figure 5. A comparison of the 700 nm backscattering data plotted as the binned transmitted (top panel, left) and the internally recorded data (top panel, middle) from Navis Float 0039 deployed off Hawaii in 2014. The upper right plot contains both data sets. A median filter was applied to both data sets and the results plotted as the resultant filtered data (lower panel, left), the raw data minus the filtered data (lower panel, middle) and a comparison of the sharpened transmitted data plotted with the recovered data as a visual demonstration of the ability to reconstruct the raw data from the transmitted data and the number of data lines included in each bin.

From this result, where variance data is important to the result, we recommend including the number of subsamples in the binned transmitted data. While this increases the total amount of data transmitted on an

ongoing basis, the net data transmission is far lower than the total data set and allows for post processing that can recapture the variance of the original data.

#### 4. OUTLOOK

The ongoing efforts to improve the capabilities of Argo on several fronts simultaneously will provide extended data coverage for deep-ocean regions with increased accuracy, possible even beyond what ship-based observations have provided. Continued improvements in the characterization of the CTD sensor dynamics show promise for the implementation of real-time, on-board data corrections that could increase the vertical resolution of core Argo data further. The spectrum of available biogeochemical sensors will grow as technology advances. Here, the main challenge lies in maturing the currently existing sensing techniques to the point where stable measurements are possible and sampling schemes are optimized for deployment on autonomous platforms.

#### REFERENCES

- Schmitt, R.W., R.C. Millard, J.M. Toole, and W.D. Wellwood. 2005. A double-diffusive interface tank for dynamic-response studies. *Journal of Marine Research*, 63(1), 263–289, <https://doi.org/10.1357/0022240053693842>
- Zilberman N., G. Maze. 2015. Report on the Deep Argo Implementation Workshop. Hobart, May 5-7th 2015. <http://archimer.ifremer.fr/doc/00281/39238/>
- Hausfather, Z., K. Cowtan, D.C. Clarke, P. Jacobs, M. Richardson and R. Rohde. 2017. Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Science Advances*. 3:1:e1601207, DOI: 10.1126/sciadv.1601201
- Briggs, N. M.J. Perry, I. Cetinic, C. Lee, E. D'Asaro, A.M. Gray and E. Rehm., 2011. High-resolution observations of aggregate flux during a subpolar North Atlantic spring bloom. *Deep Sea Research Part I: Oceanographic Research Papers*, 58(10), pp.1031-1039, doi:10.1016/j.dsr.2011.07.007
- Antoine Poteau, H. Claustre, N. Briggs, F. D'Ortenzio, C. Schmechtig, L. M. Prieur, and Emmanuel Boss. 2016. High-resolution Bio-Argo and Argo measurements to reveal specific oceanic processes. Poster, 2016 Ocean Science Meeting, New Orleans.
- Walsh, I.D. and W.D. Gardner, 1992. Comparison of large particle camera profiles with sediment trap fluxes, *Deep-Sea Res.* (39):1817-1834.