In-situ data from two 1500+ sub-mesoscale near-Lagrangian float deployments across 14 different sensor types

Dr.Jeffrey S. Ellen
Environmental Sciences / Basic and Applied Research Division
Naval Information Warfare Center Pacific
San Diego, California, USA
jeffrey.s.ellen.civ@us.navy.mil - ORCID: 0000-0002-8604-1780

***Abstract*—Here we describe a publicly available dataset of near-Lagrangian data points collected over multiple months (Feb-Aug 2022) in the Gulf of Mexico and the Western Atlantic (primarily the Gulf Stream). As of August 24, 2022, over 500 million unique datum have been recorded at 95 million individually timestamped locations, representing over 4.4M hours of coverage; collection is ongoing. Most samples are at ~5-min intervals and ~5 km spacing, which compares favorably in cost and coverage with similar Lagrangian float deployments.**

Keywords—large scale deployment, direct ocean measurement, edge processing, low-cost, novel, data collection.

I. Introduction

These deployments were a result of the Defense Advanced Research Projects Agency (DARPA) Ocean of Things (OoT) program. The goal of this program was to deploy thousands of low-cost, environmentally friendly intelligent edge processing devices as Lagrangian drifters to collect environmental and other data [1]. Part of DARPA's mission is that "DARPA explicitly reaches for transformational change instead of incremental advances" [https://www.darpa.mil/about-us/mission], therefore these floats are not intended to be similar to existing floats, but to be 'transformational.' To that end, whereas the development and engineering of most contemporary ocean instrumentation leads to the most precise instrument possible, the philosophy of these floats and deployments is to instead prioritize quantity and low cost in order to maximize the breadth of the coverage, rather than the precision.

This paper includes a brief hardware overview of the OoT floats, a summary of the available data, examples of available data, details on data publication and availability, and concludes with a discussion of how this data set compares to similar Lagrangian float deployments.

II. Lagrangian Float Hardware Overview

Design philosophy, objectives, and early prototypes are described separately[1]. Every cylindrical 3.5kg float deployed in 2022 has eight sensors (Figure 1); consisting of seven lower-powered, less-expensive sensors common to all floats and one higher-powered, more expensive 'mission sensor' per float (in a

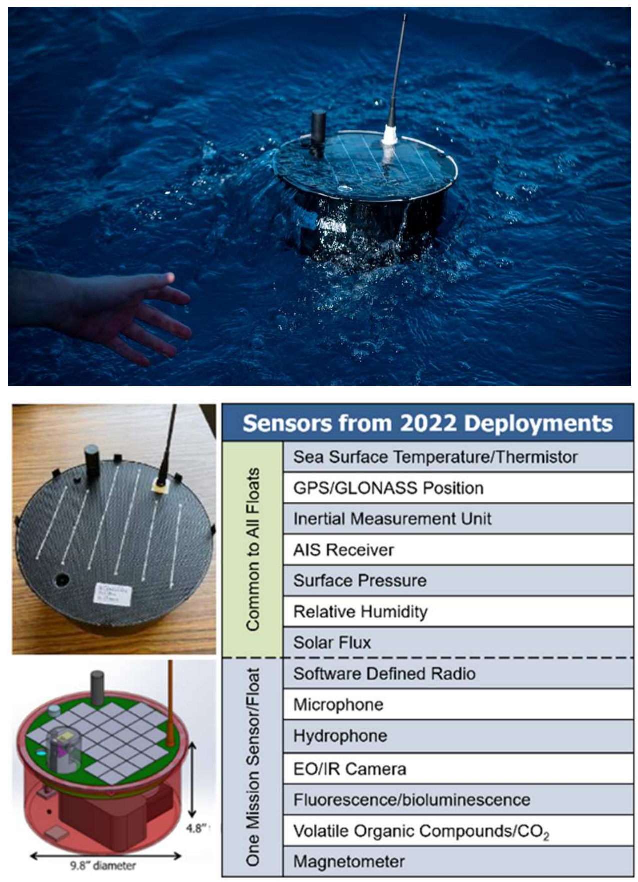

Fig. 1. Float being depolyed (top). Float top view and sketch (left) and payload summary (right). [Top photo credit: Dawson Roth/DVIDS 220213-N-KZ419-1140]

roughly equal distribution). An Iridium 9602 modem facilitates 2-way communication with the float. A solar rechargeable Li-ion cell backed up by alkaline batteries allows continuous sampling for a lifespan limited only by biofouling/float integrity failure. A Raspberry Pi single board computer provides edge processing and storage. [Raspberry Pi is a trademark of Raspberry Pi Ltd.] Float sensors sample at high rates and fidelities, which are down sampled/averaged then sent via

Funding for this effort provided by DARPA STO via BAA HR0011-20-S-0042 and BAA HR0011-18-S-0013.

Iridium Short Burst Data to shore. There is no drogue, and the freeboard is a few inches above the waterline. For an in-depth discussion of hardware specifications, see [2].

III. Data Types, Locations, and Quantity

A. Reported Float Data Types

Across the fleet's 14 different sensors, dozens of individual measurement types are reported. Most of the data collected from the seven sensors common to all floats is recorded at 5-20-min cadence, with [Position, Air Temp, Water Temp, RH, Pressure] typically reported at 5-minute intervals. For the most part, what is reported is the direct probe measurement, sometimes with less precision to save bandwidth. The specific data types available includes: Position, Air Temperature, Water Temperature, Relative Humidity, Atmospheric Pressure, Solar Power, Wave period/spectra, AIS contacts, IMU anomalies [Magnetometer, Accelerometer, Gyroscope].

Each of the floats has a single higher power "Mission Sensor" (Figure 1, lower half of table). Depending on the sensor, this data may be recorded at periodic intervals, alternatively, certain events (e.g. camera or low power acoustic detection) may trigger recordings at higher fidelities &/or power consumption. The mission sensors also typically record significantly more data than can be sent, therefore these report types are the result of edge processing algorithms to summarize the observations or report only anomalies. Specific data types available include: Optical Camera Image detections/stats, RF peaks/IQ/stats, VOC, Dedicated Magnetometer stats, Microphone Octave Band Noise, Microphone point-in-time noise, Microphone detection/anomalies, Hydrophone Octave Band Noise, Hydrophone point-in-time noise, Hydrophone detection/anomalies, Fluorometer active/passive measurements.

The power budget of the float does not allow for all collected data from all sensors to be sent with perfect reliability. There are two reasons why the sensor streams from any individual float may have variance in what is being reported. First these floats can be sent many types of parameter adjustments from shore, such as: altering duty cycle, changing what types of data are reported, or suppressing certain report types completely. The floats were actively managed during the deployments, therefore, some variance does exist, for example, sometimes floats were put into 'burst mode' collecting data at a faster-than-normal cadence. Also, all reported data is sent on best-effort basis using priority queues (which can also be configured from shore), and data over 24 hours old is rarely sent at the cost of more recently collected data. Therefore, individual floats may have gaps or changes in the type of data they reported.

For an in-depth discussion of all available data types and engineering decisions pertaining to individual report types, see [2].

B. Reported Float Data Locations and Times

Data is currently available from two OoT deployments. 1733 floats were deployed from a single ship in the Gulf of Mexico during February 10-13, 2022. 1639 floats were deployed from a single ship in the Atlantic Ocean on March 10, 2022 starting 120 km east of Norfolk, past the continental shelf break (Fig 2.). For both deployments, collection is ongoing as of August 2022, and is intended to continue throughout the lifetime of the remaining floats.

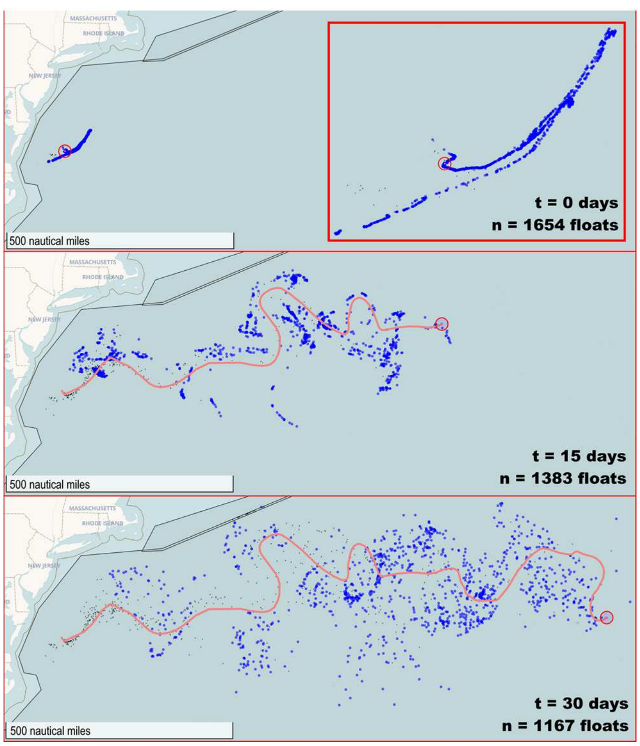

Fig. 2. March 2022 Atlantic deployment float positions at t = 0,15,30 days. Inset shows positions in detail. Red line traces the trajectory of a single float.

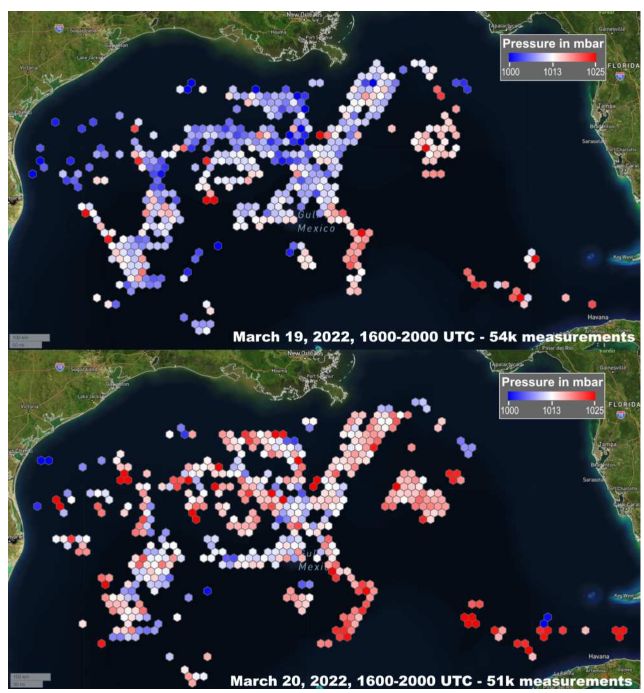

Fig. 3. Atmospheric Pressure reports from the Gulf of Mexico deployment in March 2022. Each figure consists of 4-hour averages from over 50,000 individual float reports at 5-minute intervals.

C. *Reported Float Data Rates and Examples*

Most of the common sensors take readings every 5 minutes, and that data is batched and sent every few hours, so there is some latency. These tend to be ~90% of all reported data. Additionally, other types of data are reported less frequently or when detected, and often sent immediately. For example, through August 24, float reports include: 838k AIS contacts; 4.0M VOC reports (each containing H2, ethanol, and total VOC median/max @ 5-min intervals); 3.6M VHF Maritime band activity scans, and 363k octave band noise reports (each containing 32 octaves), 3.0M daytime measurements of bioluminescence and 2.3M measurements of nighttime bioluminescence.

Figure 3 illustrates atmospheric pressure readings from 1000+ floats, most of which were reporting at 5-minute cadences. During this period, 5 weeks after the initial deployment, the floats were distributed over most of the Gulf of Mexico, with field coverage spanning over 1000 km east-west and 500 km north-south.

IV. Data Quality Control and Availability

First order data products are those that consist of compilations of directly reported data from individual floats. For first order data products, quality control (QC) flags are assigned following the IOOS QARTOD Data Flag Protocol. QC labels are assigned to flag unusual values which are likely due to faulty sensors or corrupted data. For example, pressure values outside of [970-1029] being flagged as 'suspect', outside of [950, 1033] being flagged as 'warning', and outside of [940, 1035] being flagged as 'fail'. More advanced QC is underway, such as intercalibrating adjacent floats to detect and normalize biased sensors.

First order data products are available via ERDDAP queries for the Atlantic and Gulf of Mexico deployments are available on the SECOORA website at https://erddap.secoora.org/erddap/tabledap/oot-lant-deployed-environmental.html and https://erddap.secoora.org/erddap/tabledap/oot-gomex-5-deployed-environmental.html respectively. The ERDDAP protocol facilitates both human-generated web queries as well as RESTful programmatic access, and the data can be downloaded in a variety of formats, including CSV, JSON, and NetCDF.

Second order data products include averaged, gridded, or interpolated data as well as products that are derived from inferences from the data. Examples of second order data products include inferred wind speed (derived from microphone octave band noise reports) and a Very High Frequench (VHF) Radio Activity Map (derived from software defined radio signal-to-noise ratio). These data products are under development and published to the DAPA OoT website at https://oceanofthings.darpa.mil/data as they become available.

Data are available free of charge. Please cite this paper when using these data sets.

V. Current and Projected Data Set Size

Projecting the final data set size requires evaluating the individual float lifespans. In theory, the floats are capable of operating until the rechargeable Li-ion cell no longer holds a charge. In practice, the floats are lost before that, presumably due electrical failure or water ingress due to physical housing failures. Causes are likely to include rough seas, biofouling, and prolonged UV exposure. Because the floats are not intended to

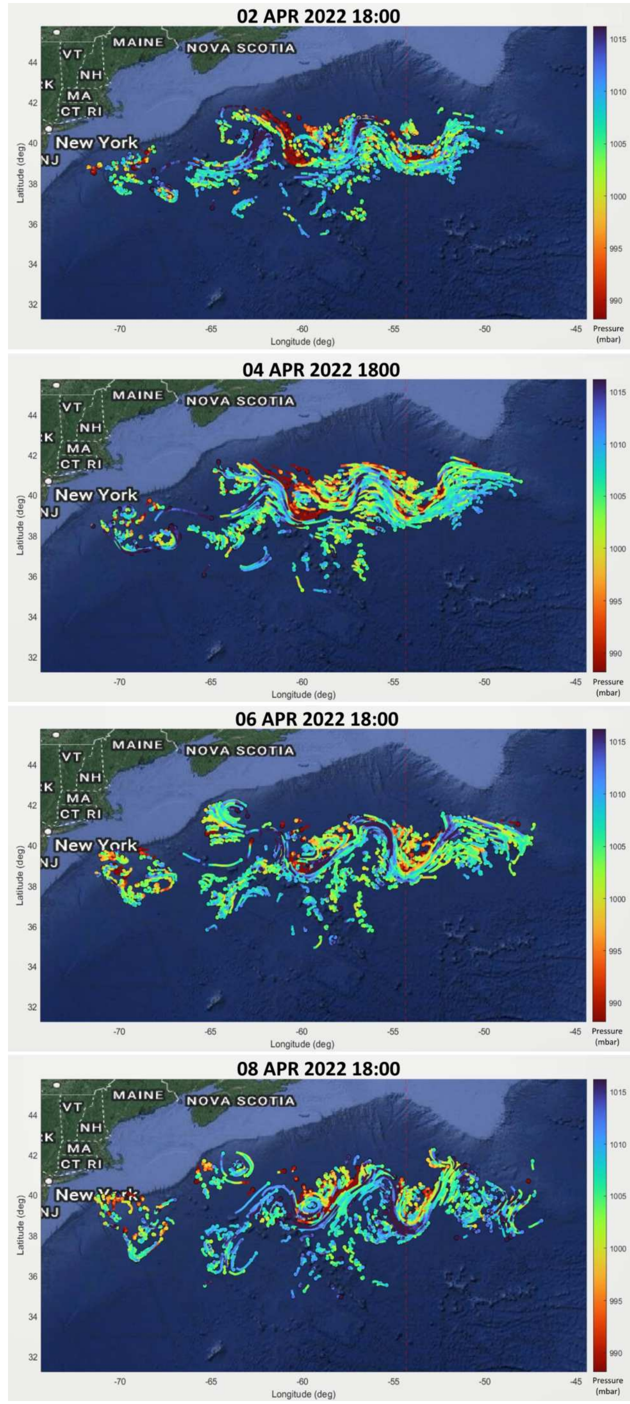

Fig. 4. Atmospheric Pressure reports from the Atlantic deployment over a week in early April 2022. Each figure consists 24 hours worth of historical reports from each float (So the top panel is showing measured pressure from the 24 hours up to and including 18:00 on April 2nd). A low pressure area is visible near (40N, 60W) in the first two panels, with floats moving cyclonically nearby. A high pressure front is visible at ~55.5W in the third and fourth panel.

be recovered, they are intentionally made with minimal amounts of plastic to attempt to minimize environmental impact [DARPA BAA HR0011-18-S-0013] and [1].

Furthermore, in order to avoid becoming a surface hazard, the floats have a scuttle mechanism designed to ensure that the float sinks quickly [DARPA BAA HR0011-18-S-0013] and [1]. This scuttle mechanism can be triggered remotely, or autonomously by the float [2]. The float is designed to scuttle upon sensing water ingress, losing GPS fix, having insufficient power available, or drifting beyond a configurable 'geofence'

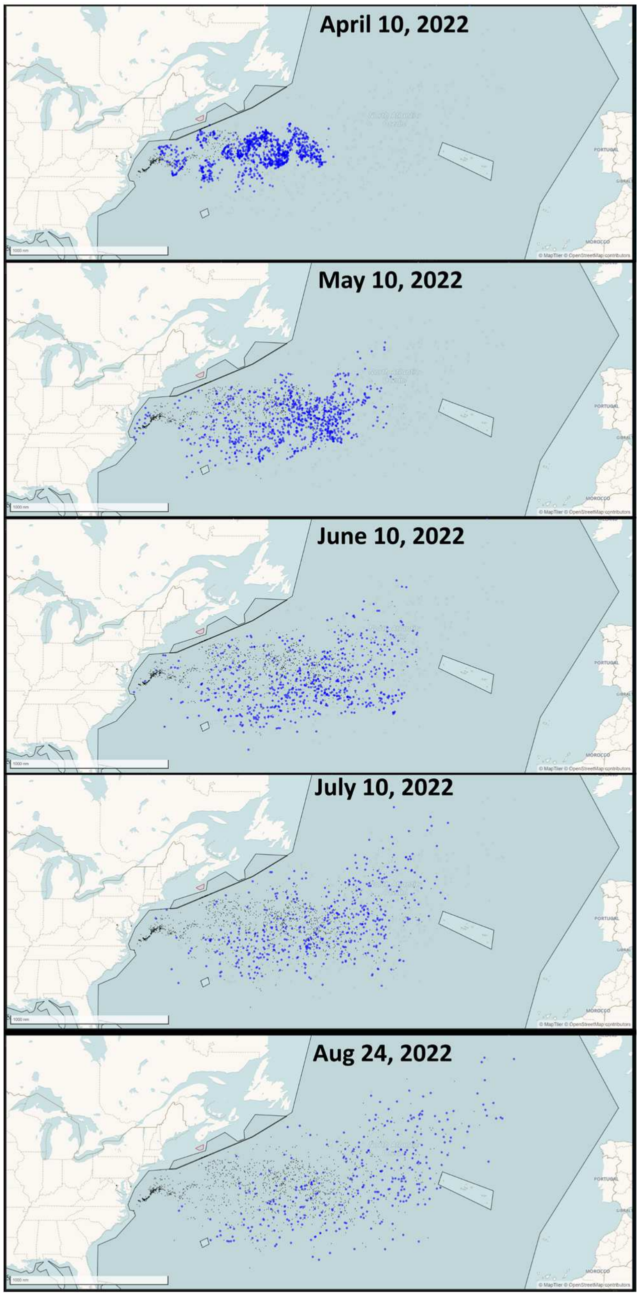

Fig. 5. Monthly current position reports from the floats of the Atlantic deployment. If the last reported position is less than 24 hours old, the float is considered 'active' and the dot is blue and slightly larger. If the last reported position is more than 24 hours old, the float is considered lost and the dot is smaller and black/grey.

[1,2]. For these OoT deployments the implemented geofence always exceeded 22 km (territorial waters).

For the Gulf of Mexico OoT deployment, average float lifespan has significantly been affected by our geofence configurations. As of Aug 24, 2022, there were only 9 active floats remaining. From deployment through 00:00Z on August 24th, a total of 21.4M reports have been received for atmospheric pressure at a 5-min cadence, yielding 1.78M hours of sensor coverage (12 reports per hour of coverage, Table 1). There are a similar number of air temperature, water temperature, humidity, etc. A total of 37.5M unique timestamped locations have been reported because some measurements, especially for mission sensor observations, come out of cycle.

For our Atlantic OoT deployment, the geofence has had minimal impact on the float field. As of Aug 24, 2022, there are 315 floats remaining. From deployment through 00:00Z on August 24th, a total of 21.4M reports have been received for atmospheric pressure at a 5-min cadence, yielding 2.61M hours of sensor coverage (12 reports per hour of coverage, Table 1). A total of 57.9M unique timestamped locations have been reported. Snapshots of the float locations are shown in Figure 5.

TABLE I. OCEAN OF THINGS DEPLOYMENT DATA SET SIZE

Location	Data Set Size Metrics			
	Unique Timestamped Locations	**Pressure Reports**	**Floats Reporting**	**Sensor-Hours**
Gulf Of Mexico	37.5M	21.4M	1733	1.78M
Atlantic	57.9M	31.3M	1639	2.61M
Combined	**95.4M**	**52.7M**	**3372**	**4.39M**

Overall, the most numerous observations are those at ~5-minute intervals, with ~52M each. Those comprise the majority of the 500M observations. For trajectory analysis, there are some discontinuities in every float's reporting history. Cumulative number of reports for each float are shown in units of sensor-hours of coverage in Figure 7. Individual float coverage length is relevant primarily for calculations requiring a high density of floats (which must occur earlier in the deployment), or calculations requiring very long individual float reports, such as a month-long GPS trajectory. For example, 1125 floats provided coverage of 720 hours or more (~1 month) in the Atlantic deployment, and 1070 floats provided coverage of 720 hours or more in the Gulf of Mexico deployment.

For the remaining 315 Atlantic floats, we conservatively expect an additional 9M total float reports over the next 7 months for each of the common sensor types. The 9M total is based on 20% float attrition per month. The 20% attrition rate is an average of past few months of non-geofence scuttle, and is more conservative than the projection in the next section (however, storms will cause variance). This will result in ~40M total reports (an additional 750k sensor-hours) for these data types over the 12-month timespan.

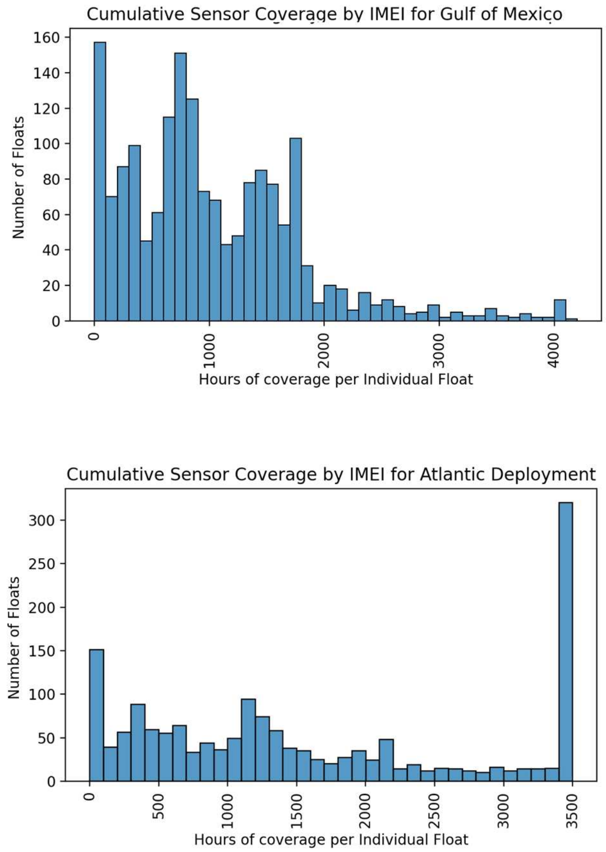

Fig. 7. Distributions of the sensor coverage provided by individual floats, per deployment. The Gulf of Mexico deployment has more losses earlier primarily due to floats scuttling due to geofence excusion, especially early in the deployment. The 3400-3500 hour bin is large for the Atlantic deployment because it includes all 315 currently active floats.

VI. Comparison to Existing Floats and Deployments

A. Comparison to Four similar float deployments: Argo, SVP, Sofar, and LASER.

Lagrangian float observations fill a gap in spatio-temporal observation of the ocean between satellites and fixed buoys and moorings (like those that are part of NOAA's National Data Buoy Center), which are especially costly in the pelagic. Ships are capable of high-resolution pelagic measurements, but at high costs and are not capable of supporting long residence times. In considering two bellwether, ongoing observation systems (ARGO and SVP) the data sets from these OoT deployments complements these systems by making more frequent and more closely spaced observations than those systems, as well as making observations on novel sensor types at far less cost (although less precise). Similarly, when considering two experimental, newer deployments of lower cost lagrangian/ocean-going floats (LASER and Sofar), the data from these OoT deployments are similarly paced and spaced to those deployments, but again are more numerous overall, as well as more diverse in types of sensors deployed.

The ARGO program is currently an international effort supporting a global array of ~4000 profiling floats. The program has existed for almost two decades [3], first achieving global coverage in 2004, and scaling up to 3000 floats in 2007[4]. ARGO's focus is not on surface measurements, but on maximally precise measurements of temperature, salinity, and pressure to 2000m [https://argo.ucsd.edu/about/mission/]. More sensor types are being added, but the main additional focus is on biogeochemical (BGC) sensors (oxygen, pH, nitrate, chlorophyll, backscatter, irradiance, etc.) [5].

The Surface Velocity Program (SVP) drifters are part of NOAA's Global Drifter Program (GDP) and the internationally coordinated Global Surface Drifter Array (GSDA)[6]. SVP drifters from their inception were explicitly a low-cost device [7] focused on circulation, but also serving as a platform for additional sensors "for observing winds, salinity, temperature, ocean color, and atmospheric pressure." The focus on low cost and breadth of sensor suite is similar to the sensors deployed by OoT.

Sofar Ocean Technologies in 2019 created a distributed sensor network of "over 100 free-drifting, real-time maritime weather sensors" cwhich provided 12+ months of continuous coverage, primarily of the northern Pacific Ocean. Sofar's Lagrangian drifters are known as the "Spotter" and reported position frequently and were capable of reporting wave, wind, current, and sea surface temperature [8]. For the time period under consideration, the data from these two OoT deployments is at a higher spatial density with many more sensor types, although Sofar advertised a total of 1000 floats deployed as of December, 2020 [9].

Overall, the nature of the data is similar in temporal and spatial scale to the Lagrangian Submesoscale Experiment (LASER) deployment[10]. The drifters were called "Consortium for Advanced Research on Transport of Hydrocarbon in the Environment" (CARTHE) drifters, and designed to be biodegradeable. The CARTHE drifters were deployed in batches during Jan-Feb 2016 in the northern Gulf of Mexico which cumulatively recorded over 10M positions at

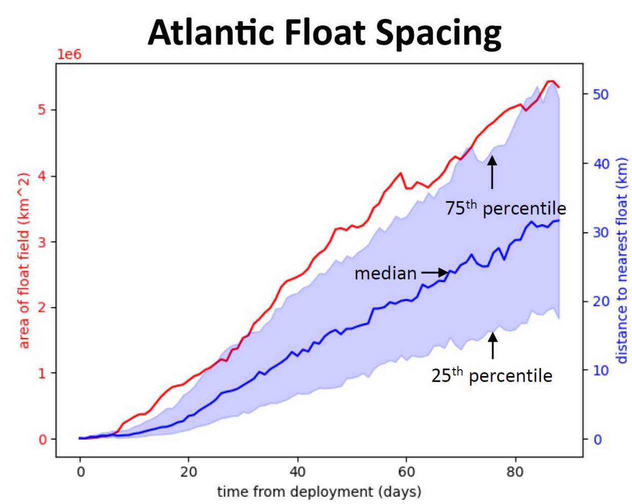

Fig. 6. Float spacing and coverage from the Atlantic deployment through early June 2022. At the end of 90 days, the median float spacing is 30km, and the area of the float field is over 5.5 million km2.[2]

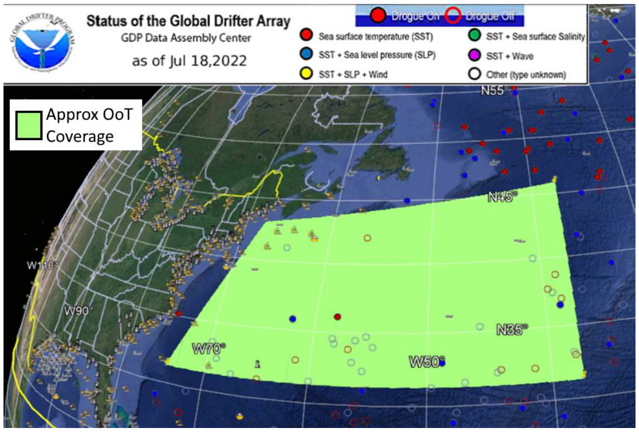

Fig. 8. Comparing coverage area of the GDP to the OoT Atlantic deployment. Hundreds of OoT floats were recording data inside the green box during the month of June and July 2022, compared to less than 25 GDP sensors covering the same region.

five-minute intervals over a three-month span [10]. However, GPS was the only sensor onboard these CARTHE drifters.

B. *Comparison of Rate of Attrition*

Excluding geofence scuttles, the combined float attrition rate for these two deployments at 7 days is 12%, the rate at 30 days is 25%, and the rate at 60 days is 45%. This projects to a failure rate of 83% at 365 days (see [2] for additional analysis). This rate of attrition is similar to NOAA's GDP, which annually replaces ~77% of its floats. ["About 1000 drifters are deployed each year by the GDP and its international partners to maintain the global 5° x 5° gridded array of ~1300 drifters." https://www.aoml.noaa.gov/phod/gdp/faq.php]. Recent ARGO floats have had a lifespan of 5-7 years, and the BGC ARGO floats are intended to have a lifespan of 4 years[5], so their rate of attrition is ~15-25%. The LASER deployment did not mention data collected after 3 months, presumably the total is minimal. No information on lifespan or attrition is reported by Sofar. Note that Sofar and OoT floats have rechargeable batteries with solar panels, and can potentially last many years.

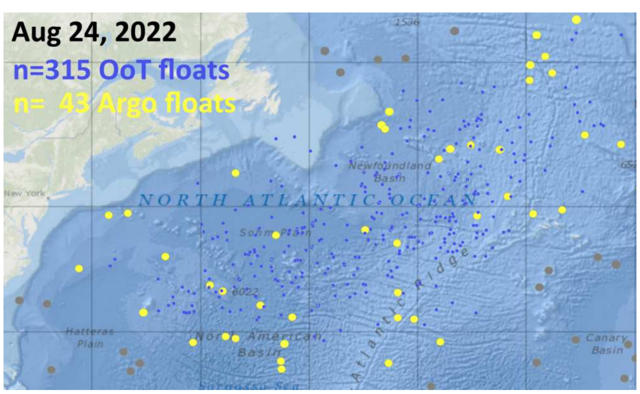

Fig. 9. Comparing coverage of the Argo network in the vicinity of the OoT Atlantic deployment on August 24. The coverage ratio is ~7:1 of OoT floats to Argo floats. Earlier in the deployment, with 5x the amount of OoT floats in less than half of the geographic area, the ratio would have been closer to 70:1. [Image from https://argovis.colorado.edu/]

C. *Comparison of Float Cost*

At a unit cost of <$500-$1000 at scale [DARPA BAA HR0011-20-S-0042] and [1], the floats from these OoT deployments cost ~1 order of magnitude less than NOAA's SVP Global Drifters ($3k-$10k, depending on sensor package), [https://www.aoml.noaa.gov/phod/gdp/faq.php] and [6], and ~2 orders of magnitude less than Argo floats ($20k-$150k) [https://argo.ucsd.edu/about/] and [4] or BGC floats ($100k) [5]. The CARTHE drifters had low-cost as a primary design goal, but no cost was provided in the paper. As of August 2022, the Spotter is available from Sofar for $5k [https://www.sofarocean.com/products/spotter].

D. *Comparison of Observation pacing and spacing.*

ARGO, GDP, and Sofar all deploy floats from multiple vessels throughout the world in order to achieve global coverage. The LASER deployment eventually occupied most of the Gulf of Mexico, and the same thing happened with the OoT Gulf of Mexico deployment (e.g. Figure 3). The OoT Atlantic deployment is a hybrid. They began with sub-mesoscale spacing, and eventually the floats settle into a distribution straddling sub-mesoscale to mesoscale, with median float spacing of 30km after 90 days (Figure 6).

As a consequence, the float density for the OoT deployments was much greater than ARGO or GDP, as shown in figures 8 and 9. It is unplanned coincidence that the ratio of relative unit cost of OoT floats to SVP drifters floats is approximately 10-1, which is the same ratio as their geographic coverage in the Atlantic deployment. Similarly, the cost ratio of OoT floats to ARGO floats is ~50-1, and the geographic coverage is in a similar ratio.

VII. Conclusion

The DARPA OoT program set out to create a dataset unlike anything currently available in Oceanography, and by quantity the deployments are a success. Community scrutiny will reveal the true utility. The datasets are being made available to facilitate investigations complimentary to existing data (e.g. using ARGO floats or moorings to calibrate OoT floats) as well as facilitate investigations not previously possible.

Contemporary terrestrial meteorology (as opposed to oceanographic meteorology) leverages both expensive, exquisite instruments such as doppler radars as well as large networks of backyard anemometers and rain gauges. Similarly, detection and measuring of tectonic movement or traffic patterns used to be the exclusive domain of precise, highly specialized equipment like seismometers or pneumatic road tubes, but both are now augmented by relatively noisy, imprecise measurements from cell phones [11] and [12]. Even within the domain of satellites, one of the most expensive of all instruments, the concept of a smaller/cheaper approach (CubeSat) is decades old [13]. These OoT floats and the data they collected is less expensive per float and per datum than any other comparable program, and higher volume (when considering either spatial resolution, temporal resolution, or both). By being in line with these trends from other domains, an extension of that same logic would conclude that the data sets described in this paper as well as the concept of the OoT-caliber float are therefore worthy of further exploration.

ACKNOWLEDGMENTS

Leadership and vision for the OoT program was provided by DARPA including program management by John Waterston with assistance from many, including Jason Rhea and Ethan Madison. NIWC Pacific assisted with concept, implemented data collection, and conducted validation of early prototype floats, facilitated by guidance from Pat Earley and Leslie Bolick and invaluable assistance from Jessica Carilli. The Navy Research Lab detachment at John C. Stennis Space Center spearheaded verification of correct float operation and execution of both 2022 deployments, all of which was organized by Blake Landry. The floats were designed and built by a team at Palo Alto Research Center, including Julie Bert, Eric Cocker, and Frank Torres. Data Analysis and figures were created by numerous government and contractor personnel, including teams at Leidos, Axiom Data Science, Metron, and Geometric Data Analytics. These four teams were led by Andrew Wiggins, Kyle Wilcox, Bryan Osborn, and Paul Bendich, respectively. Valuable contributions were made by many other individuals.

REFERENCES

[1] J. Waterston, J. Rhea, S. Peterson, L. Bolick, J. Ayers, and J. Ellen, “Ocean of things: Affordable maritime sensors with scalable analysis,” IEEE OCEANS 2019-Marseille, pp. 1-6, June 2019. doi: 10.1109/OCEANSE.2019.8867398

[2] E. Cocker, J. Bert, F. Torres, M. Shreve, J. Kalb, J. Lee, et al., “Low-Cost, Intelligent Drifter Fleet for Large-Scale, Distributed Ocean Observation.” IEEE OCEANS 2022-Hampton Roads submission (in press).

[3] S. Wilson, "Launching the Argo armada." Oceanus-Woods Hole Mass.- 42.1, pp. 17-19, 2000.

[4] D. Roemmich, G. C. Johnson, S. Riser, R. Davis, J. Gilson, W. B. Owens, et al., “The Argo Program: Observing the global ocean with profiling floats.” Oceanography, 22(2), pp.34-43, 2009.

[5] D. Roemmich, M. H. Alford, H. Claustre, K. Johnson, B. King, J. Moum, et al., “On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array.” Frontiers in Marine Science 6, pp. 439, Aug 2019. doi: 10.3389/fmars.2019.00439

[6] L. R. Centurioni, J. Turton, R. Lumpkin, L. Braasch, G. Brassington, Y. Chao, et. al, "Global in situ observations of essential climate and ocean variables at the air–sea interface," Frontiers in Marine Science 6, pp. 419, Aug 2019. doi: 10.3389/fmars.2019.00419

[7] P. Niiler, “The world ocean surface circulation”, International Geophysics, Volume 77, pp.193-204, 2001. doi: 10.1016/S0074-6142(01)80119-4

[8] I. A. Houghton, P. B. Smit, D. Clark, C. Dunning, A Fisher, N. J. Nidzieko, et al., "Performance Statistics of a Real-Time Pacific Ocean Weather Sensor Network." Journal of Atmospheric and Oceanic Technology, Volume 38, Issue 5, p.1047-1058. May 2021.

[9] A. Freedman, “Company deploys sensors over half the world’s oceans to improve weather forecasts,” Washington Post: Weather, December 18, 2020.

[10] G. Novelli, C. M. Guigand, C. Cousin, E. H. Ryan, N. J. M. Laxague, H. Dai, et al. “A Biodegradable Surface Drifter for Ocean Sampling on a Massive Scale.” Journal of Atmospheric and Oceanic Technology, Volume 34, Issue 11, p. 2509-2532. Nov 2017. doi: 10.1175/JTECH-D-17-0055.1

[11] S. Dashti, J. Reilly, J. D. Bray, A. Bayen, S. Glaser, E. Mari, et al. “iShake: Using personal devices to deliver rapid semi-qualitative earthquake shaking information.” GeoEngineering Report. 2011.

[12] G. Rose. "Mobile phones as traffic probes: practices, prospects and issues." Transport Reviews, Volume 26.3, pp. 275-291, May 2006.

[13] H. Heidt, J. Puig-Suari, A. Moore, S. Nakasuka, and R Twiggs, R. “CubeSat: A new generation of picosatellite for education and industry low-cost space experimentation.” 14th Annual AIAA/USU Conference on Small Satellites. 2000.