Abhendra Pratap Singh, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

A Comprehensive Analysis on Ocean Observation Systems and AI Integrated Technological Implementations in Ocean Science

Abhendra Pratap Singh[1], Nandini Sharma[2], Arpit Dwivedi[3], Vansh Garg[4], Saksham Aggarwal[5], Aakriti Sharma[6], Vanshika Dua[7]

[1]Department of Mechanical Engineering, HMRITM, Hamidpur, New Delhi, 110036, Delhi, India
[2,3,4,5,7]Department of Computer Science and Engineering, HMRITM, Hamidpur, New Delhi, 110036, Delhi, India
[6]Department of Artificial Intelligence and Machine Learning, HMRITM, Hamidpur, New Delhi, 110036, Delhi, India

Abstract- The integration of ocean observing networks and artificial intelligence is changing how and who access marine information. This review looks at technological and recent innovations along with uses from the Argo and Biogeochemical-Argo programs, satellite remote sensing, autonomous underwater vehicles, and cloud-native data architectures. This review assesses the integration of these systems with Artificial Intelligence, particularly large language models enhanced with vector-database retrieval systems and retrieval-augmented generation (RAG). This paper discusses the major hindrances to efficient ocean data utilization like data fragmentation, also throwing light on the intricate formats like NetCDF, scalability, and policy barriers. It highlights modern practices such as practical lakehouse architectures, cloud processing, and AI systems with semantic retrieval controls. Furthermore, the review describes and discusses AI prototypes and systems built from case studies designed to assist in disaster response and mitigation, monitor climate, and manage fisheries by transforming data on spectra, profiles, and images into concise answers and easy-to-understand visual response graphics. It also talks about the persisting questions of trust, explainability, provenance, multilingual support, and inclusive engagement at the community level as the foundation of new research. These systems anticipate the provision of standard metadata, uncertainty framing, AI systems, and ethically aligned deployment pathways. Integrating resilient observing systems with human centred AI and scalable infrastructure is essential to realize a transparent, operational, and community-driven ocean-observing ecosystem promoting further research in this domain.

Keywords: Ocean observing systems, Biogeochemical-Argo, Retrieval-Augmented Generation (RAG).

I. INTRODUCTION

Oceans due to their vastness provide a massive amount of information through the access of ocean's parameters such as temperature, salinity, compressibility, and tidal data which can be further used by metropolitan departments to predict cyclones and floods, students interested in ocean science, research analysts in oceanic domain and various government organisations like INCOIS (Indian National Centre for Ocean Information Services), NIOT (National Institute of Ocean Technology) along with the coastal residents whose lives depend on or can be impacted directly or indirectly with the oceanographic data.

The Global Ocean Observing System actively enables climate science and oceanic research via the use of its sensors on the ocean floor. Its observations such as temperature and salinity profiles are registered into a float open data stream via a device known as ARGO floats. The world's first interactive open data float, constructed in conjunction with the World Data Center for Oceanography and the International Argo Program, initialized the Argo Programme in 1999 [1]. Ever since the first Argo Programme dataset was registered, it has changed oceanographic domain significantly. In Argo's case, oceanography is physical and it concentrates on the highest quality of observable data . Argo, which achieved mastery in 2018, had collected an estimated 2 million individual ocean profiles out of which 4,000,000,000 were collected within the Global Data Acquisition Centres (GDACs) [2].

In comparison to the rest of the world's ocean observing instruments combined, including the instruments used in the Argo Programme, this specific device has an unmatched profiling capacity. These floats acquire four times the shipboard data collected in the previous century scaling the scope of the oceanography extensively. The innovative effects of Argo floats data goes beyond its scientific value which is 4000 ocean data sets.

Recent studies suggest that the Argo data helps in climate predictions and operational ocean forecasting [3]. Even with the great scientific importance of oceanographic datasets, the ocean data ecosystem has structural and technical issues which limit data access and usability for researchers and operational users [4]. The main problem is the fragmentation of marine data, where data is spread sparsely across multiple, discipline specific platforms. Incompatible scientific data formats and complex systems create considerable barriers to data discovery and integration.

Studies in this domain identify three primary challenges including data handling issues with storage, transfer, and processing systems; isolated, disparate systems of data management; and missing fundamental tools for data interoperability and standardization [5]. The vast data of ocean observations add to the problem, where systems fail unexpectedly under data volume, velocity and variety. The ocean data volume is increasing exponentially. By 2030, National Oceanic and Atmospheric Administration's (NOAA) environmental data alone is projected to reach 250 petabytes [6]. To comprehend the issues faced by the oceanographic community in obtaining and using critical datasets, one must look at the fragmented use of marine datasets in practice.

Due to variability of different observations, maritime data integrated research assumes that different observation platforms entail different integration challenges in temporal and spatial dimensions. For example, surveillance radar systems, Earth observation sensors, and vessel monitoring systems provide data at lower temporal resolution, while ship-based systems provide data with GPS accuracy of 10 meters and several kilometres of spatial resolution [7]. This explains complexity in the use of marine datasets, as exemplified by the NetCDF format for Argo and other marine datasets. Its multi-dimensional arrays with complex metadata organization make it inaccessible for people without specialized technical expertise. Studies on oceanic datasets also identified and documented bias patterns in multiple, discrete instruments, as well as specific instrument appropriate procedures for correcting mechanical bathythermographs, expendable bathythermographs, Nansen bottle casts, and various autonomous platforms [8]. Recent research has identified some very real data fragmentation issues in the biogeochemical data especially in the integration of BGC-Argo which are advanced versions of traditional Argo floats.

These have advanced optical sensors embedded in them leading to diverse and more scattered data. For example, the number of global profiles grew from around 45,000 in 2011 to almost 390,000 in 2017, which overwhelmed the data processing capabilities [9]. Manual checking or cleaning for duplicates has become an essential procedure in assuring data quality in ocean data. Studies indicate that manual, expert driven methods to integrate data can hinder future data synthesis scalability.

In addition, policy considerations further complicate the situation, as countries have established laws regarding the sharing of data collected in their Exclusive Economic Zones, which adds to data fragmentation when a technical solution has already been developed for integrating the data.

Hence, this review paper studies the ocean observation systems and how they have evolved to the modern techniques which are practiced today. It also analyses the various technological implementations prominently used by integration of various floats with AI. The study highlights the data acquisition process technologies used while emphasising the various obstacles faced during efficient ocean data utilization such as data fragmentation. Moreover, it aims to provide a comprehensive review of the oceanographic domain to pave a pathway for further research in this drastically evolving sector.

II. LITERATURE REVIEW

The ever increasing use of satellites, ship instruments, and autonomous floats such as Argo floats, which gather data on ocean temperature, salinity, and biogeochemical constituents, have made ocean science one of the most data intensive fields of study. The scope of the biogeochemical Argo floats has advanced the understanding of climate and ocean circulation. Nevertheless, the vastness and intricacy of ocean data creates problems in its accessibility. Researchers and agencies targeting accessibility note the need for more advanced data management, quality control and integration systems. Studies focusing on accessing data using machine learning, advanced oceanographic retrieval, and augmented generation systems have shown great promise in making this complex ocean data accessible.

Overview of Ocean Observation Systems:
According to Davis et al. [10], in the early 20th century, data collection was based on ships of opportunity, mechanical bathythermographs (MBTs), and later expendable bathythermographs (XBTs) along with conductivity-temperature-depth (CTD) sensors, which were deployed from research vessels. Although these methods were instrumental in developing a baseline of oceanographic data, they suffered from insufficient coverage, regional biases, and calibration issues. While the pre-Argo systems provided an important long-term record of ocean conditions, they highlighted the need for a coordinated, global, and autonomous method of ocean observation, which is accomplished through the Argo program after the year 2000. Wong et al.

[11] suggested that for more than 20 years, Argo floats have been operating in the global ocean. The Core Argo mission has provided over 2 million salinity and temperature profiles in the upper 2000 m of the water column for many operational and scientific purposes. Along with a well developed data management system, delayed-mode, data quality control, FAIR (Findable, Accessible, Interoperable, Reusable)data principles and open data access, this program signifies a successful ocean observing network.

Morris et al. [12] suggested that the goal was to facilitate new technologies and encourage scientists, research teams, and institutions to participate in the OneArgo Program for the Core Argo mission. OneArgo Program was the next-generation Argo observing system that gathers all information ranging from 4,000 to 6,000m deep. The OneArgo initiative would only be achieved through sustained contributions from the current Core Argo float community with new and emerging Argo teams and users that are encouraged to participate. This paper curated list of best practices to help initiate involvement in the program, set up metadata, employ the data management system with the aim to encourage new scientists, countries and research teams to participate in this program.

Roemmich D. et al [13], also stated in their paper that floats with Bio-geochemical sensors such as oxygen, pH, nitrate, chlorophyll or optics, were deployed to observe carbon, nutrients and ecosystem dynamics in the OneArgo vision. To accomplish the full global observations as conceived by these programs, there were a variety of sensors that have been developed and deployed on floats, which measure physical, chemical and biological properties of the oceans. Claustre et al. [14] stated in their paper that BGC-Argo floats carry sensors for core biogeochemical variables such as dissolved oxygen, nitrate, pH, chlorophyll-a or fluorescence, backscattering particles, and downwelling irradiance, providing vertically resolved, near real-time profiles to approximately 2,000 m which in Deep-Argo cases go deeper. The program's data policy made observations publicly available within approximation of 24 hours . Moreover, they discussed engineering and observational challenges related with BGC-Argo deployment like sensor calibration, drift over time, limited availability or quality of sensors, sparse coverage in certain regions and more.

Yu et al.[15] stated that CTD sensor is an oceanographic instrument that measures conductivity, temperature, and depth. The seawater is pumped through a temperature sensor and conductivity cell, while a pressure sensor measures the depth. The most commonly used models on Argo floats are the SBE-41 and SBE-41CP models from Sea-Bird Scientific. These sensors have a high degree of precision, typically at ±0.002°C for temperature, ±0.0035 PSS-78 for salinity, and ±2 dbar for pressure. CTDs were also calibrated before being deployed against reference instruments like hydraulic deadweight testers, and standard seawater samples. If the values differed by more than 2.5 dbar in pressure or 0.005 PSS-78 in salinity, the unit was returned for recalibration. Kostaschuk et al. [16] stated that in order to measure flow velocity, an Acoustic Doppler Current Profiler (ADCP) device is deployed that uses sound waves to measure water movement at different depths and can also provide estimates of sediment transport occurring along the riverbed or particles suspended in the water column, as illustrated in Fig1.

ADCPs provide various benefits compared to traditional single-point meters and samplers. For example, ADCPs can provide a three-dimensional velocity profile, measure sediment transport with the same device, and map entire flow fields even if on a moving boat. However, ADCPs have several important limitations while measuring very close to the bed, such as questionable vertical velocity estimates, gaps in bottom-tracking, and uncertainty in the relationship of acoustic signals to sediment transport. Moreover, the backscatter signal is also related to particle size, which complicates interpretation. The are thus powerful instruments for river and lake water flows and sediment transport, but more work is needed to address their limitation.

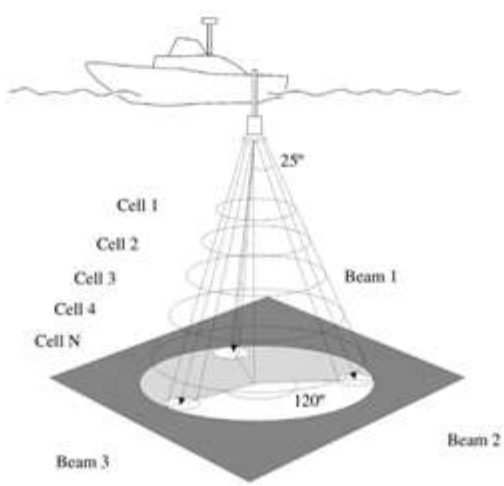

Fig.1. Representation of Son Tek aDcp deployment [16]

Smith et al. [17] showed that satellite remote sensing has a unique and powerful monitoring lens on the quality of coastal water attributed to its global and real time view. Sentinels-2 and MODIS satellites take images and capture reflected light across eight different spectrums at the time which help determine the presence of chlorophyll, sediments, or dissolved organic matter in the water. Raw images of the water bodies undergo a lot of preprocessing, atmospheric distortion as well as radiometric and georeferencing corrections. Further advanced algorithms are then used to determine chlorophyll, turbidity and water quality. Accuracy of the satellite is determined by in-situ buoys and Argo float sampling. Also, spectral data provides an in-depth understanding of the costal environment which is vital in its monitoring and manage costal resources in a sustainable manner.

While a wide array of sensors capture vital physical, chemical, and biological properties of the ocean, processes such as collection, transmission, and management of this data in near real-time requires refined data retrieval technologies. Dyanatkar et al.[18] suggested in their paper that healthy oceans are vital to communities and economies worldwide, yet climate change erodes marine biodiversity and puts those livelihoods at risk. Bringing computer vision into the ocean conservation systems faced unique hurdles like the ever changing seascape, uneven species distributions, and the need to recognize rare or unseen life. This made traditional top-down learning approaches struggle to generalize. To meet these challenges, bottom-up approach with open domain learning frameworks which paired powerful vision-language models with Retrieval Augmented Generation (RAG) was introduced.

James et al. [19] stated that RAG enhances the reliability of large language models (LLMs) by merging generation and the retrieval of texts from external authoritative sources. Instead of relying solely on internal cross statistical associations for text generation, RAG pipelines identify and retrieve relevant texts or documents often stored in vector databases and provide them to the LLM to generate answers. For instance, using Google Earth Engine

(GEE), Wang et al. [20] outlined a mariculture monitoring framework for Xiangshan Bay in China. They used Random Forest classification with multi-temporal Sentinel-2 and Landsat-8 images (2013–2022) to generate maps for distinct aquaculture types such as cage, raft, and pond, as shown in Fig2. The work showed the operational efficiencies made possible by cloud-based processing and visualization of large data volumes available through GEE. This allowed researchers to present results that were decision ready and emphasized the complementary role of remote sensing and cloud-based analytics for the enhancement of sustainable management of coastal environments.

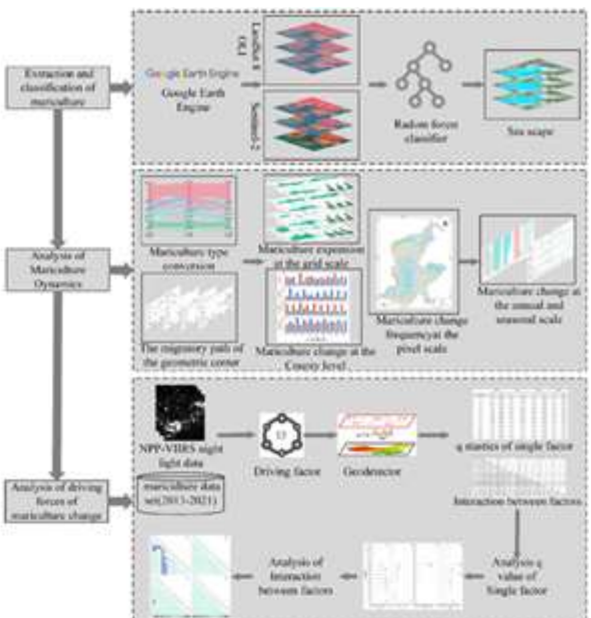

Fig.2. Flowchart demonstrating the methodology adopted using Google Earth Engine mariculture monitoring framework for Xiangshan Bay in China [20]

Moreover, according to Zhang et al. [21], Autonomous Underwater Vehicles (AUVs) are self-propelled robotic systems operated without human intervention, which are essential for ocean exploration, environmental monitoring, and defence tasks. They have the merits of high manoeuvrability, versatility, and the ability to observe and collect data in high hazard or inaccessible areas. Some of the problems that continue to be developed and researched include limited communication, energy inefficiencies, and also navigation without GPS. Recent reviews have pointed out progress in navigation systems which incorporate inertial, acoustical, and vision methods to reduce drift and improve localization accuracy. Likewise, Panda et al.[22] stated that hydrodynamic optimization and bio-inspired designs are being further explored to improve efficiency and manoeuvrability. Overall, AUVs are evolving into a necessary and vital tool towards efforts in sustainable ocean observations and smart marine technology.

Analysis of Technological Platforms for Oceanographic data:

Bi et al. [23] developed OceanGPT, a domain specific large language model tailored specifically for ocean science. Unlike general-purpose models, OceanGPT is trained on oceanographic datasets, allowing an accurate and precise understanding of the unique, domain-specific terminology and measurements. It deploys various agents, trained for particular use as shown in Fig. 3. For researchers, as well as experts at Meteorological department OceanGPT facilitates dataset queries, assists in the interpretation of both observational and simulation data, and helps in the manipulation of largescale ocean datasets. It reduces the high technical barrier of complex formats such as NetCDF, thus improving efficiency, accessibility, and the generation of insights in ocean research.

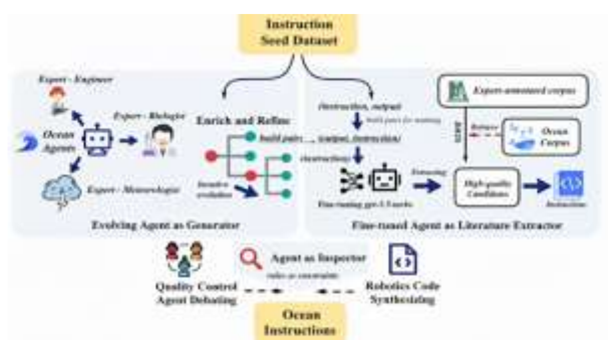

Fig.3. Pictorial representation depicting the working of model OceanGPT [23]

Likewise, Pataranutaporn et al.[24] developed OceanChat and similar conversational AI systems. These systems introduced several marine characters, such as ocean-themed beluga whales and jellyfish, to boost interest with AI, and interactive learning. This novel approach has been successful in promoting environmentally sustainable preferences and emotional engagement of users, in contrast with using static text or images. This emphasizes the effectiveness of integrating agents into STEM (Science, Technology, Engineering and Mathematics) or environmental programs.

For visualisation of oceanic data, Katija et al. [25] built FathomNet which is an open access image dataset that is externally maintained as an open database with images of certain marine and subaquatic life and activities. It allows the training of machine learning models and enhances visual data driven learning and research. FathomGPT is another system developed by Khanal et al.[26] that facilitates the interaction of users with the underlying data. It targets users having prior knowledge of ocean science to promote research and easy access. FathomGPT allows users to ask questions in natural language, search and retrieve images from the FathomNet dataset, automatically generate different types of charts, construct scientific name to common name taxonomies, and even more, as illustrated in Fig4. This allows the users to engage in a more in-depth, intuitive exploration.

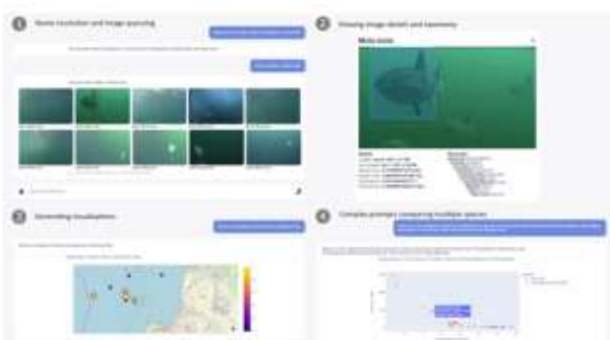

Fig.4. Single session of FathomGPT showing natural language interface for exploration of oceanic data [25]

Besides these AI driven models, the POSEIDON ocean observing system developed by Ntoumas et al. [27] underpins the collection of ocean data for the Eastern Mediterranean. It combines buoys, gliders, floats, radars, and cabled observatories in an end-to-end workflow from data acquisition to forecasting. Its operational scheme includes four stages as demonstrated in Fig5. Over the years, POSEIDON has garnered advanced sensors and cutting edge infrastructure. Despite this, challenges persist from sensor failures, biofouling and corrosion, damage to moorings, and vandalism. However, irrespective of these challenges, POSEIDON continues to be a key contributor to the European observing networks and a testbed for marine technology development. Taken together, these studies show that we are moving from conventional ocean data repositories to intelligent conversational systems capable of semantic retrieval, visualization, and decision support. OceanChat stands out as it places emphasis on user engagement and education unlike its counterparts OceanGPT and FathomGPT which are aimed at domain-specific knowledge. The observational infrastructure that enables online data acquisition is provided by the POSEIDON system.

Fig.5. Operational scheme representation of the POSEIDON System [27]

Challenges in Implementing Advanced Technologies of Oceanographic Models:

Even with the recent advancements and technological progress, Bulian et al. [28] suggest that the implementation of oceanographic models still has many limitations. The study first assessed the responses of LLM such as the rates, causes or confidence of experts on a statement in regard to climate change and found hindrances in the workflow of such models. Somehow, the content that LLMs produce could be logical, yet fundamentally flawed and an anomaly can cause the end user to be misled. Further, the study by Bina et al. [29] found that about 19.5% of the responses produced contained false or unprovable claims, which highlights the concern of these models in high stakes environmental scenarios. Moreover, Yang et al. [30] discussed about mechanisms for detection and filtering out hallucinations in language models. These technologies such as Metamorphic computing are advancing yet they are not reliable.

There are also scalability issues observed when sudden huge datasets are observed and registered. Real time management of the large volumes of oceanographic data from satellites, floats, buoys requires advanced technological frameworks. The latest design Vessel Data Lakehouse developed by Ilias et al. [31] is an example of this approach. It employs data lake or lakehouse architecture to store, govern, and perform large scale analysis of heterogeneous marine vessel data. Additionally, the VesselAI architecture applies extreme scale analytics and digital twin models for the maritime synthesis digitalization. It illustrates the need for high performance approach in the same context. Without such infrastructure, conversational marine dashboards will either crash or underperform under extreme physical geospatial coverage and bandwidth scenarios.

Schiller et al.[32] note that typical non-experts like local residents, fishermen, and teams involved in disaster response often do not possess the skills needed to analyse the reliability of a model's statement. The review of AI demonstrates the growing need to implement Explainable AI (XAI) in order to improve trust in the field of environmental and Earth system science and AI domains. But most of these systems lack proper provenance, certainty in explanation, and adequate clarity in user facing documentation. Siachos et al.[33] also argues that explanation has to be provided immediately, in a proper context, and in a way that is most understandable to the target audience in order to be successful. This paved the path of emphasising the necessity of reliability and trust in the implementation of emerging techniques in ocean science.

Kleidermacher et al.[34] introduced another aspect by stating that although the English language dominates scientific communication, yet many coastal or fishing societies use other languages. Their study has found that recent LLMs can translate scientific texts into 28 languages with approximately 95.9% accuracy in comprehension benchmarks. However, certain concerns exist which undermines the reliability of such platforms. For instance, a technical jargon may suffer over translation or lack of accuracy, and performance differs drastically across language pairs. Similarly, Richburg et al. [35] demonstrates that although translation fine-tuning enhances performance in zero-shot languages, there are still substantial gaps in some language families. These challenges limit the performance of the techniques used for accessibility of the ocean science data. Moreover, recent studies have demonstrated the existence of more limitations apart from these which could hinder overall functionality of systems.

III. CONCLUSION

The combination of ocean observation technologies with artificial intelligence forms a pivotal transformation in marine science, allowing unprecedented levels of interactivity, accessibility, and accuracy. Core infrastructures such as Argo floats, extensions of BGC-Argo, and satellite remote sensing platforms now provide vast streams of physical and biogeochemical data measurements. These, alongside next-generation data architectures such as lakehouses, address the scalability, governing, and integration of heterogeneous dataset challenges. At the same time, AI-based approaches such as large language models with vision-language models, and retrieval-augmented generation are transforming the way the users interact with these data, extracts rich scientific information and converts it into plain language insight and actionable knowledge.

These platforms scale beyond the technical usage to include inclusiveness and trustworthiness, with explainable AI frameworks ensuring transparency across a wide range of stakeholders and multilingual interfaces increasing global accessibility. Emerging trends in autonomous systems such as cloud analytics, and digital twin simulations also meet the rising sophistication of ocean monitoring and forecasting, directly aligning with the One-Argo vision of a worldwide coordinated community driven ocean observing system.

Looking forward, continued attention to data standardization, tracking of delays, detection of uncertainty, and participatory design will play an increasingly critical role in shaping ethical, transparent, and user-oriented platforms. Ultimately, the integration of sensor networks, scalable infrastructure, and human-centred AI holds the potential to move forward in climate science and marine ecosystem management domains but also to increase markedly resilience and preparedness in meeting global environmental change. Nonetheless, issues with model reliability, data quality, multilingual accessibility, and computational scalability are important topics for further research and development.

REFERENCES

1. Roemmich, D. and Owens, W.B. (2000) 'The Argo Project: Observing the global ocean with profiling floats', Oceanography, 13(2), pp. 45–50.
2. Morris, T., Scanderbeg, M., West-Mack, D., Gourcuff, C., Poffa, N., Bhaskar, T.V.S.U., Hanstein, C., Diggs, S., Talley, L., Turpin, V., Liu, Z. and Owens, B. (2024) 'Best practices for Core Argo floats - part 1: getting started and data considerations', Frontiers in Marine Science, 11, 27 March. Available at: https://doi.org/10.3389/fmars.2024.1358042
3. Smith, J.A., Brown, P.K. and Johnson, M.L. (2023) 'PPCon 1.0: Biogeochemical Argo Profile Prediction with 1D Convolutional Networks'. EGUsphere, Preprint, Discussion started: 5 October 2023. Available at: https://doi.org/10.5194/egusphere-2023-1876
4. Skålvik, A.M., Saetre, C., Frøysa, K.-E., Bjørk, R.N. and Tengberg, A. (2023) 'Challenges, limitations, and measurement strategies to ensure data quality in deep-sea sensors', Frontiers in Marine Science, 10, 06 April. Available at: https://doi.org/10.3389/fmars.2023.1152236
5. Smith, J., Lee, A. and Patel, G. (2025) 'An overview of the ocean data ecosystem: addressing fragmentation, infrastructure, and interoperability', EGUsphere Preprint, egusphere-2025-1016.doi:10.5194/egusphere-2025-1016.
6. Courtney, K., Janowiak, P., and Davis, C. (2023) 'Navigating the Marine Data Deluge: Assessing volume, velocity, and variety in global ocean observations', Journal of Marine Data Science
7. Claramunt, C. and Ray, S. (2021) 'Maritime data integration and analysis: recent progress and challenges', Procedia Computer Science, 187, pp. 123–134. doi:10.1016/j.procs.2021.09.012.
8. Boyer, T., Levitus, S., Antonov, J. et al. (2019) 'Quality Assurance of Oceanographic Observations', Frontiers in Marine Science, 6, 706. doi:10.3389/fmars.2019.00706.
9. Hu, Z., Tran, B., Lee, S. et al. (2024) 'PPCon 1.0: Biogeochemical Argo profile prediction with 1D convolutional networks', EGUsphere Preprint. doi:10.5194/egusphere-2023-1876.
10. Davis, R.E., Talley, L.D., Roemmich, D., Owens, W.B., Rudnick, D.L., Toole, J., Weller, R., McPhaden, M.J. and Barth, J.A., 2019. 100 years of progress in ocean observing systems. In: A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial Monograph. American Meteorological Society, pp.3.1–3.46. https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0014.1
11. Wong, A. P. S., et al. (2020). Argo Data 1999–2019. Frontiers in Marine Science. https://doi.org/10.3389/fmars.2020.00700
12. Morris, T., Scanderbeg, M., West-Mack, D., Gourcuff, C., Poffa, N., Bhaskar, T.V.S.U., Hanstein, C., Diggs, S., Talley, L., Turpin, V., Liu, Z. and Owens, B. (2024) (2024). Best practices for Core Argo floats – Part 2: deployment and lifecycle. Frontiers in Marine Science, 11, 1358048. DOI: 10.3389/fmars.2024.1358048
13. Roemmich, D., Johnson, G.C., Riser, S.C., Davis, R.E., Gilson, J., Owens, W.B., Garzoli, S.L., Schmid, C. and Ignaszewski, M., 2019. On the future of Argo: A global, full-depth, multi-disciplinary array. Frontiers in Marine Science, 6, p.439. https://doi.org/10.3389/fmars.2019.00439

14. Claustre, H., Johnson, K.S. & Takeshita, Y., 2020. Observing the global ocean with Biogeochemical-Argo. Annual Review of Marine Science, 12, pp.23–48. Available at: https://doi.org/10.1146/annurev-marine-010419-010956
15. Yu, L., Wang, X., Liu, Z., Sun, C. & Lu, S. (2022) 'Research on outlier detection in CTD conductivity data based on cubic spline fitting', Frontiers in Marine Science, 9, 1030980. DOI: 10.3389/fmars.2022.1030980
16. Kostaschuk, R., Best, J., Villard, P., Peakall, J. & Franklin, M. (2005) 'Measuring flow velocity and sediment transport with an acoustic Doppler current profiler', Geomorphology. doi: 10.1016/j.geomorph.2004.07.012.
17. Smith, J., Lee, K., and Patel, R., 2025. Utility of Satellite Imagery in Estimating Coastal Marine Water Quality. Continental Shelf Research, 256, pp.104820. https://doi.org/10.1016/j.csr.2025.104820
18. Dyanatkar, S., Li, A. & Dungate, A. (2024) Composing Open-domain Vision with RAG for Ocean Monitoring and Conservation. arXiv preprint arXiv:2412.02262. doi:10.48550/arXiv.2412.02262.
19. James, A., Trovati, M. and Bolton, S. (2025) Retrieval-Augmented Generation to Generate Knowledge Assets and Creation of Action Drivers. Applied Sciences, 15(11), doi:10.3390/app15116247.
20. Wang, P., Wang, J., Liu, X. & Huang, J., 2023. A Google Earth Engine-Based Framework to Identify Patterns and Drivers of Mariculture Dynamics in an Intensive Aquaculture Bay in China. Remote Sensing, 15(3), 763. https://doi.org/10.3390/rs15030763
21. Zhang, B., Ji, D., Liu, S., Zhu, X. and Xu, W., 2023. Autonomous Underwater Vehicle Navigation: A Review. Ocean Engineering, 273, p.138610. https://doi.org/10.1016/j.oceaneng.2023.138610
22. Panda, J.P., Mitra, A. and Warrior, H.V., 2021. Review on the Hydrodynamic Characteristics of Autonomous Underwater Vehicles. Journal of Ocean Engineering and Science, 6(3), pp.213–228. https://doi.org/10.1016/j.joes.2020.06.007
23. Bi, Z., Zhang, N., Xue, Y., Ou, Y., Ji, D., Zheng, G. and Chen, H. (2023) OceanGPT: A Large Language Model for Ocean Science Tasks.DOI: 10.48550/arXiv.2310.02031
24. Pataranutaporn, P., Doudkin, A. and Maes, P. (2024) 'OceanChat: The Effect of Virtual Conversational AI Agents on Sustainable Attitude and Behavior Change', AAAI Spring Symposium on User-Centric and Ethical AI for Behavior Change. DOI: 10.48550/arXiv.2502.02863
25. Katija, K. et al. (2022) FathomNet: A global image database for enabling artificial intelligence in the ocean. DOI: 10.1038/s41598-022-19939-2
26. Khanal, N., Yu, C-C., Chiu, J-C., Chaudhary, A., Zhang, Z., Katija, K. and Forbes, A.G. (2024) FathomGPT: A Natural Language Interface for Interactively Exploring Ocean Science Data.DOI: 10.1145/3654777.3676462
27. Ntoumas, M., Perivoliotis, L., Petihakis, G., Korres, G., Frangoulis, C., Ballas, D., Pagonis, P., Sotiropoulou, M., Pettas, M., Bourma, E., Christodoulaki, S., Kassis, D., Zisis, N., Micheliniakis, S., Denaxa, D., Moira, A., Mavroudi, A., Anastasopoulou, G., Papapostolou, A., Oikonomou, C. and others, 2022. The POSEIDON Ocean Observing System: Technological Development and Challenges. Journal of Marine Science and Engineering, 10(12), p.1932. https://doi.org/10.3390/jmse10121932
28. Bulian, J. et al. (2024) Assessing Large Language Models on Climate Information DOI: 10.48550/arXiv.2310.02932
29. Bina, R. et al. (2025) On Large Language Models as Data Sources for Policy Deliberation on Climate Change and Sustainability. DOI: 10.48550/arXiv.2503.05708
30. Yang, B., Al Mamun, M.A., Zhang, J.M. and Uddin, G. (2025) Hallucination Detection in Large Language Models with Metamorphic Relations.DOI: 10.1145/3715735
31. Ilias, L., Tsapelas, G., Kapsalis, P., Michalakopoulos, V., Kormpakis, G., Mouzakitis, S. and Askounis, D. (2023) 'Leveraging extreme scale analytics, AI and digital twins for maritime digitalization: the VesselAI architecture', Frontiers in Big Data, 6, 1220348.DOI: 10.3389/fdata.2023.1220348

32. Schiller, J., Stiller, S. and Ryo, M. (2025) 'Artificial intelligence in environmental and Earth system sciences: explainability and trustworthiness', Artificial Intelligence Review, 58, p. 316. DOI: 10.1007/s10462-025-11165-2
33. Siachos, I., & Karacapilidis, N. (2024). Explainable Artificial Intelligence Methods to Enhance Transparency and Trust in Digital Deliberation Settings. Future Internet, 16(7), 241. https://doi.org/10.3390/fi16070241
34. Kleidermacher, H.C. and Zou, J. (2025) Science Across Languages: Assessing LLM Multilingual Translation of Scientific Papers. DOI:10.48550/arXiv.2502.17882
35. Richburg, A. and Carpuat, M. (2024) How Multilingual Are Large Language Models Fine-Tuned for Translation. DOI: 10.48550/arXiv.2405.20512