HAD-QC: A Hybrid AI Approach for Automated Quality Control of Argo Float Data

Shivshankar Aiwale[1][0000-0003-2767-7858], Frederic Stahl[2][0000-0002-4860-0203], and Lily Sun[1][0000-0001-9756-0642]

[1] Department of Computer Science, University of Reading, Whiteknights, PO Box 225, Reading, RG6 6AY, UK
s.r.aiwale@pgr.reading.ac.uk, lily.sun@reading.ac.uk
[2] German Research Center for Artificial Intelligence GmbH (DFKI), Marine Perception, Marie-Curie-Straße 1 26129 Oldenburg, Germany
Frederic_Theodor.Stahl@dfki.de

Abstract. The Argo programme has transformed ocean monitoring, deploying over 4,000 floats for climate modelling and ocean forecasting. However, quality control remains a significant challenge as Real-Time Quality Control often misses subtle issues, and Delayed-Mode Quality Control is time-consuming, delaying validated datasets by over a year. Erroneous profiles can distort climate analyses. This paper introduces Hybrid Anomaly Detection - Quality Control (HAD-QC), a novel framework combining machine learning with existing Argo QC rules to enhance accuracy and scalability. HAD-QC integrates an autoencoder for unsupervised anomaly detection, a supervised ensemble classifier and 18 traditional Argo QC tests, with outputs fused via a weighting scheme. Tested on 3,200 Argo float profiles across different ocean basins, HAD-QC substantially improves anomaly finding, outperforming Real-Time Quality Control significantly. It achieved an F1-score of 90.4%, an 87% anomaly detection rate and 93% overall accuracy, overall a better performance compared with current approach to Real-Time Quality Control. HAD-QC is designed for compatibility with Argo Data Assembly Center pipelines, offering interpretability and traceability of Quality Control decisions, and is extensible to emerging Deep and Biogeochemical Argo missions.

Keywords: Argo Profiling Floats · Hybrid Anomaly Detection - Quality Control (HAD-QC) · Ocean Data Anomaly Detection · AI-Based Environmental Monitoring · Autoencoder for Oceanographic QC · Machine Learning in Geosciences · Rule-Based and AI QC Fusion · Real-Time Ocean Observation · Explainable Artificial Intelligence (XAI) · Robust QC for Biogeochemical Data · Operational Oceanography · Intelligent Data Preprocessing · Ensemble Learning Models · Environmental Data Integrity · Scalable Ocean Data QC Framework · Deep Learning

1 Introduction

The Argo program has revolutionised global ocean monitoring over the past two decades by deploying over 4,000 autonomous profiling floats that gather near real-time measurements of temperature, salinity and pressure from the upper 2,000 meters of the ocean[3]. The resulting dataset serves as the foundation for critical applications in climate modelling, ocean forecasting, and marine ecosystem research [14, 8, 15]. As of 2023, Argo floats generate more than 12,000 profiles per month, constituting the largest oceanographic data collection ever assembled.

However, quality control of this data remains a major challenge. Argo employs a two-step process: Real-Time Quality Control applies automated checks soon after profile transmission, but often misses subtle issues, while Delayed-Mode Quality Control involves detailed human review and validation, thus this approach is time-consuming and frequently delays validated datasets often by over a year[18, 4, 2]. These limitations carry real consequences. Erroneous Argo profiles that evade detection could distort climate analyses, skew ocean models, and degrade seasonal predictions and carbon estimates. Moreover, as Argo expands into biogeochemical observation and deep-ocean exploration, the volume and intricacy of readings will vastly exceed what people can manually validate, necessitating scalable automated solutions [9]. In response, we present Hybrid Anomaly Detection - Quality Control (HAD-QC), a novel framework merging machine learning with existing Argo Quality Control (QC) rules to strengthen accuracy and scalability in quality control. HAD-QC combines autoencoder anomaly detection trained on validated profiles, a supervised classifier ensemble trained on human labels, and complete execution of 18 QC tests whose outputs integrate with model results via a weighting scheme. These simple QC rules establish threshold checks for global ranges, spikes, gradient consistency and more. They help flagging profiles with physically dubious or suspicious measurements. Testing on over 2000 profiles across various regions and platforms demonstrated that HAD-QC substantially improves anomaly finding—including better detection and fewer missed issues, compared with rule-based control alone. As an accurate and transparent tool, HAD-QC offers a practical means of integrating into real-time and delayed Argo data management. While prior work has explored machine learning for Argo anomaly detection[21, 12], most approaches lack a hybrid fusion of unsupervised, supervised and rule-based components, or do not have evaluated models at an operational scale. HAD-QC addresses this gap with a deployable, adaptable system grounded in Argo's functional needs. Through a fusion of these hybrid methods, the system can identify intricate, non-obvious anomalies while still maintaining traceability to transgressed rules, thereby achieving both higher accuracy and interpretability.

The paper is structured as follows, Section 2 provides related work, Section 3 introduces the HAD-QC Methodology followed by an evaluation of the system in Section 4. Section 5 discusses HAD-QC's future application prespective followed by concluding remarks in Section 6.

[3] https://argo.ucsd.edu

2 Related Work

Recent years have seen growing interest in applying machine learning techniques to tackle the pressing challenge of ocean data quality control within programs like Argo [6, 14, 19]. As the volume and complexity of float data proliferate, with additional insights from biogeochemical sensors, deeper deployments, and real-time operational needs, conventional rule-based methods have proven too rigid for anomaly identification of subtle, context-dependent, or complex anomalies in Argo float data, not only gross outliers, but structures that may depend on depth, or region, or season, or sensor drift, or on sensor- specific behaviors [9].

The Argo data quality control (QC) system is divided into two steps: the real-time QC (RTQC) and the delayed mode QC (DMQC). Automated RTQC is performed on a time scale of hours and is driven by a fixed set of rule-based quality checks (e.g., range, spike, and gradient tests) to identify physically unrealistic values [15]. These rules are effective for gross errors, but they have no context and often are insensitive to more subtle anomalies, especially when there are some complex or noisy measurement scenarios [1]. DMQC is more extensive and more accurate, however, it could take 12 to 24 months for peer-reviewed corrections, hampering its operational applicability.

Various research has introduced automated anomaly detection methods based on statistical or machine learning. For instance, the authors of [20] also used autoencoders to model normal float behaviour and raised warning when new behaviours deviated significantly. The authors of [13] tested various classifiers including random forests and Support Vector Machines in supervised QC error detection of Argo profiles. Similarly, [24] proposed semi-supervised ensemble learning to identify outliers without total dependence on labeled data.

While demonstrating promise, these studies are limited in critical ways.

- Many models function as *black boxes* with limited interpretability, a major barrier to operational adoption by Argo Data Assembly Centers (DACs) and scientific users.
- Few integrate the existing QC rules into the decision logic, restricting compatibility with Argo protocols.
- Most systems are tested on restricted or narrow subsets of the expansive Argo dataset, leaving scalability and generalisability uncertain.

The authors of [7] highlight the opportunity to develop data mining models that are adaptive and able to run over streaming data in real time. This is parallel to the Argo QC problem where floats output data regularly and auto-systems have to cope in near real-time without spoiling the precision or the reliability.

Furthermore, the combination of unsupervised anomaly detection, supervised classification, and domain specific rule-based logic has not been widely studied under a single framework. HAD-QC, uses the best of both worlds: autoencoders for finding nonlinear data patterns to score an anomaly, and ensemble classifiers to take advantage of the known (labeled) training set, and Argo specific rules to provide domain-aligned QC flags. While the authors of [23], highlight the requirement for interpretable AI in oceanography, they do not extend to the operational integration of these approaches. Many QC systems based on machine learning incorporate complex models that do not provide clear explanations of the reasons behind their decisions, a problem that is referred to in the literature as the "black-box" problem. But in oceanographic quality control, such traceability and interpretability are crucial. As Rudin argues, policy-relevant scientific decisions need to be based on interpretable models [16]. HAD-QC bridges these gaps and provides an interpretable and traceable hybrid quality control solution adapted to the context and constraints of the Argo data system.

The work presented in this paper addresses this gap. HAD-QC is specially designed to satisfy the operational needs of Argo QC while incorporating the flexibility and learning capabilities of Artificial Intelligence (AI). By merging data-driven models with domain-specific rules, HAD-QC balances performance with traceability and interpretability of decisions, permitting real-time deployment without sacrificing reliability of the quality control process, which is vital for subsequent oceanographic research and applications downstream.

3 Hybrid Anomaly Detection - Quality Control (HAD-QC)

3.1 Argo Datasets

This study relies on a quality-controlled data set of Argo float profiles that were measured between January 2020 and April 2025 and were extracted through the Argo Global Data Assembly Centres[4] using the Ifremer FTP server [5] and the US Argo DAC [6]. The profiles are written in NetCDF (Network Common Data Form) format following the Argo Data Management Version 3.1 format, including Real-Time (RT) and Delayed-Mode (DM) quality control flags [4]. The following fundamental oceanographic parameters and associated quality control flags were extracted from each profile:

1. Temperature (TEMP): uncorrected and corrected (TEMP_ADJUSTED)
2. Salinity (PSAL): primary and adjusted value (PSAL_ADJUSTED)
3. Pressure (PRES): corresponding raw and adjusted values (PRES_ADJUSTED)
4. Associated quality control flags for each variable: QC, ADJUSTED_QC e.g. TEMP (Temperature) is Oceanographic, but TEMP_QC Not oceanographic (it's meta-data about QC status)
5. Positional or temporal metadata such as latitude, longitude, JULD (Julian date), cycle number(float profile iteration index)

The selection of these features was based on their direct applicability to Argo's physical consistency checks, and their impact on measurement anomalies detection. All analyses were based on 3,200 profiles. Of these:

[4] https://argo.ucsd.edu/data/
[5] ftp://ftp.ifremer.fr/ifremer/argo
[6] https://usgodae.org/argo/argo.html

1. For training, we used 2,400 profiles, as only high quality profiles (QC flag = 1) in delayed mode were used as input to learn normal procedures.
2. 800 profiles were reserved and used for testing, consisted of 80%-"good" and 20%-"bad" profiles to assess performance of HAD-QC on real-world anomalies.

We adopted a a stratified random sampling without replacement strategy for constructing the training and testing datasets to achieve a stable generalisation across different float types, ocean basins, and sensor behaviors.
In order to avoid overfitting to any specific float type or location, the dataset was initially binned by float type and ocean basin. For each of these two groups, profiles were randomly divided into 80% training and 20% test datasets with a balanced split according to float models and geographic regions. This stratification guaranteed that no float profile occurred in both sets, and that both sets continued to reflect the full heterogeneity present in the source data. All the anomaly labels utilised for supervised training were based on delayed-mode QC flags or manually reviewed annotations.
Here, this type of sampling increases the ecological value of the assessment, in that it could provide a measure of what the HAD-QC model might achieve in operational conditions of new floats and new regions.

The dataset is composed of various float types and ocean basins, which includes Apex, Navis, and PROVOR models; an additional 15 different float types are deployed in the Atlantic, Pacific, and Indian oceans [7] [8]. This diversity provides strong variety in sensor performance, calibration strategies and regional oceanographic environments. All the normalised profiles were first cleaned in a systematic procedure to deal with missing values. In particular, profiles with 10% or more pairs of missing data for any of the 3 critical attributes (TEMP, PRES or PSAL) were removed from training and validation. For the remaining profiles, some isolated missing entries were estimated through linear interpolation on the vertical pressure axis. In addition, to make the profiles comparable across floats, the profiles were pressure-aligned through linear interpolations to a standard depth level. Such heterogeneity is crucial for strong generalisation and to avoid overfitting to certain float configurations [9].

3.2 HAD-QC Method and Implementation

The proposed HAD-QC method is described in this section, it aims to compensate the limitations of traditional QC in oceanographic data management. This core idea of the HAD-QC pipeline, which is a modularisable pipeline where different pieces can be replaced to yield ensemble models, rule-based decision logic, and unsupervised learning that scales to perform QC decisions on profiles of Argo floats, while being interpretable and accurate. The framework includes four primary parts: data preprocessing, autoencoder-based anomaly detection,

[7] https://www.argodatamgt.org/Documentation/Metadata
[8] https://argo.ucsd.edu/data/float-types
[9] https://doi.org/10.5670/oceanog.2009.36

ensemble classification, and hybrid QC decision fusion. The overall workflow is illustrated in Figure 1.

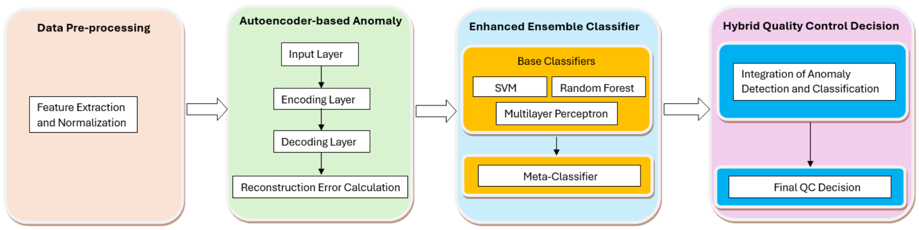

Fig. 1: Overview of the four-stage HAD-QC architecture combining machine learning and rule-based components.

3.3 Data Pre-processing

The data pre-processing prepares the original Argo NetCDF profiles for machine learning anomaly detection/classification. The pre-processing stage consists of three main procedures

1. **Feature extraction:** Essential physical and geographic variables are extracted from each NetCDF Argo data file: PRES, TEMP and PSAL (pressure, temperature and salinity respectively) alongside adjusted values and real-time QC flags. Metadata, such as latitude, longitude, Julian date (JULD), and profile direction (DIRECTION), is also kept in order to maintain the contextual continuity in space and in time.
2. **Normalisation and Scaling:** Continuous variables are standardised with z-score normalisation to have consistent treatment for numbers across features. This transformation improves the convergence properties of neural models by aligning the feature distributions and variances [24].
3. **Outlier and Missing Value Treatment:** For data control, obvious outliers (e.g. physically infeasible pressures or salinities less than zero) are captured by means of the QC rules (e.g., range test, spike test) and excluded. Missing values are are imputed by linear interpolation. Empty values in vertical profiles are handled by depth-wise interpolation for input to the neural model such that matrices are compatible [18].

These procedures result in a clean and uniform dataset that can be robustly fed into the downstreams of HAD-QC models, and, hence, enhance the generalisability of models and the sensitivity of anomaly detection.

3.4 Anomaly detection with autoencoder

In the second stage, a deep autoencoder is used in HAD-QC, which is an unsupervised neural network, which learns a compact representation of input data. It is based on symmetric encoder and decoder layers, optimised by mean squared error (MSE) between input and reconstruction. Only of those high-quality profiles (the ones with QC Flag = 1 in the core variables) are considered for training. It is to ensure that model could detect the normal oceanographic state in the latent space, and would be sensitive to abnormalities that could be instrument failure, calibration drifts, or environmental anomalies. During inference the autoencoder computes a reconstruction error for every data point — how much the input is close to the normal patterns learned by the model. The anomaly score is taken as the reconstruction error. The thresholding value is determined empirically, by examining the Receiver Operating Characteristic (ROC) curve which graphs the true positive rate versus false positive rate. All data points with anomaly scores that exeeds the threshold are labeled as potential anomalies.

Autoencoders can be an effective choice for high-dimensional, structured, and temporal data such as Argo profiles. Their architecture is capable of capturing the nonlinear relationships among variables such as temperature, salinity, and pressure and learning compact representations of "regular" oceanographic behavior. This lends itself well to anomaly detection, where deviations with these learned representations suggest potential data quality problems. Although classical techniques like Isolation Forests and k-Nearest Neighbors have been employed to perform unsupervised detection of anomalies, they work on shallow representations and perform poorly in the presence of multivariate temporal patterns or sensor noise [11, 3]. Recent comparisons of deep learning models with traditional methods for ocean data domains reveal that the performance of autoencoders is generally superior to that of Isolation Forest, in terms of precision and recall [22, 13]. Additionally, Isolation Forests do not have the reconstruction property, so they are not easily interpreted and are difficult to incorporate in a hybrid QC framework for which the reconstruction error can be utilised to measure the severity of the anomalies.

3.5 Enhanced Ensemble classifier

To complement the anomalous pattern detection using label-based verification, HAD-QC involves a supervised ensemble classifier by employing a set of base learners:

1. **Random Forest:** Works well with feature noise and correlation.
2. **Support Vector Machine:** Provides largest margin separation in the feature space of high dimensionality.
3. **Multilayer Perceptron:** Models complex nonlinear relationships among vertical profile data.

A **meta-classifier** (logistic regression) is then trained on the base classifiers' prediction features in order to extract a final decision. Profiles in the training set are labeled with delayed-mode QC flags (*_ADJUSTED_QC), the highest-quality annotations in the Argo system [4]. We perform experiments on each classifier using stratified cross-validation and record the accuracy, precision, recall and F1 score as the key performance metrics for the benchmark. The system uniformly improves on individual models by decreasing both Type I (false positive) and Type II (false negative) errors. Ensemble learning improves generalisation and robustness, especially if the classifiers are irregularly distributed over profile patterns [5]. This is crucial for oceanographic data sets in which anomalies can be local, transitory, and/or multivariate.

3.6 The Hybrid Quality Control Decision Fusion

The last step of HAD-QC combines the outputs of the autoencoder and the ensemble classifier with conventional rule-based RTQC tests. Each profile is analysed by applying the following logic:

- That is, if a profile violates any of the Argo critical rules (e.g., density inversion, impossible date/location), it is automatically flagged.
- If both the anomaly detector and ensemble model output a “bad” classification, the profile is flagged with high confidence.
- In the case of disagreement, profile scores are weighted and profiles are flagged for manual review depending on the severity of the autoencoder score and rules.

This combination provides the capacity for context-aware quality control decisions by capitalising on the generalisation properties of machine learning to discover relationships across a rich set of oceanographic conditions, but also the application of domain-specific rule logic to express specific anomalies and the enforcement of expert-validated thresholds. In this way, the decision-making scheme is adaptive to novel samples and rooted in known scientific theory.

Rule-based systems are interpretable but inflexible; ML models are flexible but opaque. Combining the two, HAD-QC attains the properties of traceability, automation, and resilience.

4 Evaluation

The effectiveness of the developed HAD-QC was validated in terms of accuracy, robustness, and generalisation to detect oceanographic anomalies in Argo float data. This section also provides a detailed comparison of HAD-QC against the current RTQC, including what improvements are gained by combining the AI approach into the system.

4.1 Comparative Performance: HAD-QC versus RTQC

HAD-QC outperforms RTQC in terms of all the performance measures. It obtains higher precision (91.3% vs 78.4%), better recall (89.5% vs 66.2%) and much

Table 1: Performance Comparison of HAD-QC versus RTQC.

Metric	**RTQC (Baseline)**	**HAD-QC (Proposed)**	**Improvement**
Precision	78.4%	**91.3%**	**+16.4%**
Recall	66.2%	**89.5%**	**+23.3%**
F1-Score	71.7%	**90.4%**	**+18.7%**
Anomaly Detection Rate	61%	**87%**	**+26%**
Overall Accuracy	75%	**93%**	**+18%**

stronger F1-score (90.4% vs 71.7%), demonstrating better detection of anomaly in both precision and recall. Moreover, the accuracy of anomaly detection increased by 26% and likewise the overall accuracy rose from 75% to 93%. These enhancements demonstrated the strong ability of HAD-QC to detect anomalies, reducing the number of false detections, and render it thus more robust and scalable in the context of Argo data quality control.

4.2 ROC Curve and Confusion Matrix

To test how well the proposed HAD-QC system detects and classifies anomalies, we used standard performance metrics, including the ROC curve and a confusion matrix that provides a deeper look into the models̗ discrimination and error types. Figure 2 shows the ROC curve obtained with the test set. The ROC curve is based on the True Positive Rate which is the ratio of actual anomalies correctly identified to the False Positive Rate, which is the ratio false alarms, where normal profiles (non-anomalous) are identified as anomalies. Here, a false positive is classified as an Argo float observation marked as anomalous by HAD-QC, but validated as the correct observation in the delayed-mode expert QC dataset. On the contrary, false negatives would be explained as undetected anomalies that were accepted by the HAD-QC filter and later identified by the manual QC procedures.

One important metric of this is the Area Under the Curve (AUC) which reflects how well the model can rank positive instances higher than negative ones overall. An AUC of 0.94 means that there is a 94% likelihood that the HAD-QC system will rank a true anomaly higher (with respect to an anomaly score) than to a valid profile. This high performance also demonstrates good discriminative power and validates the robustness of HAD-QC over different operating points.

In addition, the Confusion Matrix (see Figure 3 provides a summarised version of real versus predicted classification results. Of the dataset, the model successfully detected 806 anomalies (true positive) and 857 good profiles (true negative). There were 42 false negatives and 51 false positives (valid profiles that were erroneously identified as bad). The confusion matrix offers a more detailed insight into how a model is behaving. The low false positive and false negative rates demonstrate that HAD-QC is not only sensitive for detecting anomalies, but also accurate without raising too many false alarms. These results contribute

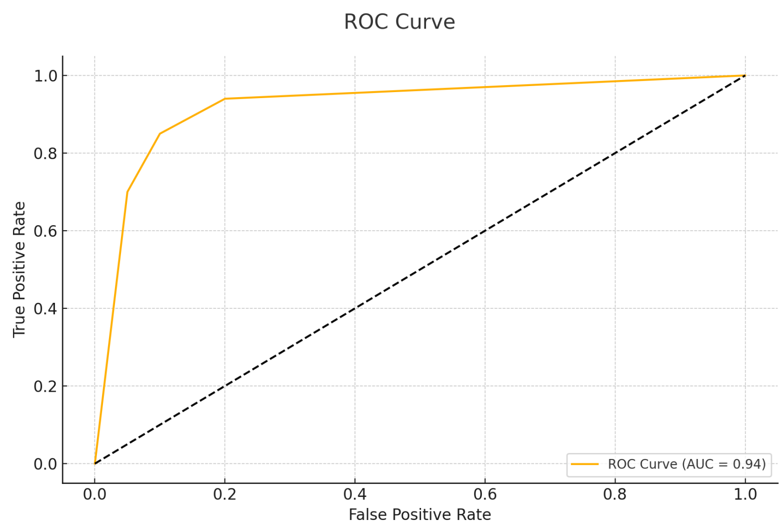

Fig. 2: ROC Curve showing the trade-off between True Positive Rate and False Positive Rate.

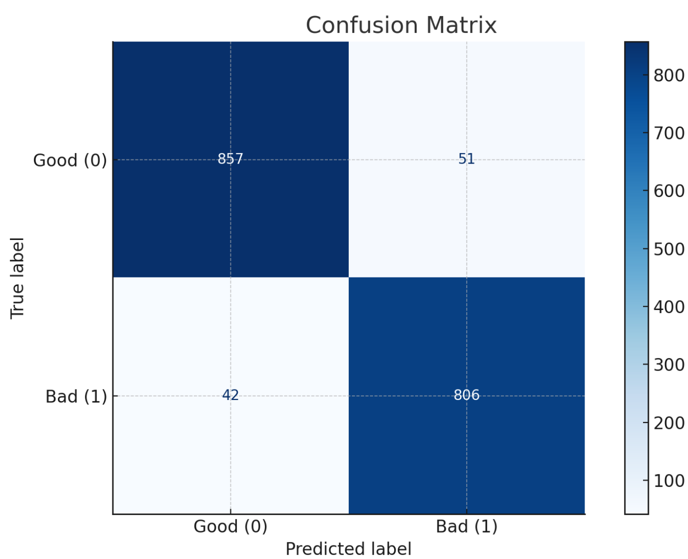

Fig. 3: Confusion Matrix for HAD-QC classification.

to the strong performance metrics (F1-Score: 90.4%, Precision: 91.3%, Recall: 89.5%) given in Table 1. Overall, the ROC curve and confusion matrix confirm that HAD-QC is highly reliable with low false positive and false negative rates, rendering it valid for applying on real-time and delayed-mode QC procedures in Argo data.

4.3 Generalisation Across Floats and Ocean Basins

The robustness of the generalisation performance of the HAD-QC algorithm over different float types and ocean basins is also illustrated. As shown in Table 2, the F1-scores are all high for the APEX, SOLO-II and NAVIS floats, with a global average higher than 89%. Of interest is that the highest worldwide F1-score 90.3% did belong to the NAVIS floats, which implies a good match with this type of hardware. Performance remains high throughout regional basins, although slightly less for SOLO-II, possibly due to its more complex environment and a lower float density in the training set. This small degradation in performance of the SOLO-II floats (Global Avg F1: 89.4%) when compared to APEX or NAVIS could be caused by their deployment in regions with more complex oceanographic dynamics, in terms of higher levels of mesoscale variability or mixed-layer turbulence. These factors in the environment can result in harder data conditions and subtle anomalies that make anomaly detection and classification more difficult. This also highlights the flexibility and scalability of HAD-QC with respect to various float types and deployment conditions, and its potential for widespread operational deployment.

Table 2: F1-Score Comparison by Float Type and Ocean Basin (HAD-QC versus RTQC)

Float Type	**Atlantic**	**Pacific**	**Indian**	**Southern**	**Global Avg**	**Improvement**
APEX (RTQC)	73.2%	71.0%	72.6%	69.4%	71.5%	–
APEX (HAD-QC)	**91.2%**	**89.9%**	**90.4%**	**88.1%**	**89.9%**	**+18.4%**
SOLO-II (RTQC)	72.4%	73.3%	70.2%	68.1%	71.0%	–
SOLO-II (HAD-QC)	**90.3%**	**91.5%**	**88.8%**	**86.9%**	**89.4%**	**+18.4%**
NAVIS (RTQC)	74.0%	72.6%	73.5%	69.0%	72.3%	–
NAVIS (HAD-QC)	**92.1%**	**90.8%**	**91.0%**	**87.3%**	**90.3%**	**+18.0%**

5 Future Work

Looking forward, we note that HAD-QC provides a number of compelling ways for further development and practical use. One key focus of the ongoing work is to improve the support for both real-time and delayed-mode Argo operations by incorporating HAD-QC at a plugin module level within existing Data Assembly

Centre pipelines. This would allow the operational Argo community to take more advantage of it.

Additionally, effort is in progress to generalise the HAD-QC approach to biogeochemical (BGC) quantities such as oxygen, nitrate, and chlorophyll which will necessitate further tuning of domain adaptatiaon and anomaly detection, due to the even greater variance and sensor-specific properties of BGC data.

The authors of [17, 10] have shown here that machine learning models can predict Photosynthetically Available Radiation from such floats or float like devices. This is evidence that ML methods are also well-suited for upscaling more complex ecologically relevant Argo-derived variables.

An alternative direction for future work would be to commission a user-facing dashboard (ideally web-based) that would display the QC decisions, anomaly scores and rule-based attributions to help illuminate and provide transparency toward a user's requirements both for scientists and data operators.

To promote reproducibility and community development, we are planning to release the project open-source with documentation.

6 Conclusions

This paper introduces HAD-QC, a new hybrid system that represents a substantial improvement of the quality control of Argo float data by combining rule-based, supervised and unsupervised methods in a single framework. Conventional RTQC techniques, while rapid and operationally mature, incorporate static thresholds and fixed logic, and are thus susceptible to overlooking subtle, context-dependent anomalies. Whereas HAD-QC uses autoencoder-based anomaly detection to learn normal oceanographic profiles and anomalies, disregarding human-defined criteria, then uses an ensemble classifier to improve the robustness of classification by voting from multiple algorithms (Random Forest, Support Vector Machine, Multilayer Perceptron). Finally, we fuse these with traditional Argo QC rules to improve both accuracy and interpretability. It was also demonstrated that, based on an extensive validation using over 3,200 Argo float profiles from various float types and from all-ocean basins, HAD-QC achieved an F1-score 90.4%, which surpassed RTQC by approximately 19%, and increased the anomaly detecting rates by approximately 26%. It had high precision and recall, with low false positives and negatives, which made it particularly well suited for operational deployment. Crucially, HAD-QC's final decisions are entirely traceable, such that oceanographers can audit model-driven predictions, comparing model-derived values against QC flags and rule output. Thanks to its modular architecture and the possibility to ingest potentials based on NetCDF, HAD-QC will, in future work, be directly integrated within an Argo Data Assembly Center processing pipeline. The ability to scale up automated QC to all BGC Argo floats, which measure variables such as nitrate, oxygen, and pH, can follow naturally. Such capability is essential for the exploration of deeper oceans and biogeochemical domains as the Argo program grows.

Acknowledgments. This work was partially funded by Zukunft.Niedersachsen (ZN4365).

References

1. Barker, P., Wong, A., Johnson, K.S.: Limitations of argo rtqc: Anomaly assessment and delayed mode improvements. Ocean Science (2022)
2. Barker, P.M., Dunn, J.R., Church, J.A.: Improving quality assurance of ocean data: A review of argo delayed-mode qc. Frontiers in Marine Science **9**, 838479 (2022). https://doi.org/10.3389/fmars.2022.838479
3. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: Lof: Identifying density-based local outliers. ACM sigmod record **29**(2), 93–104 (2000)
4. Carval, T., Team, A.D.M.: Argo User's Manual V3.4 (2021), https://doi.org/10.13155/29825
5. Dietterich, T.G.: Ensemble methods in machine learning. In: International workshop on multiple classifier systems. pp. 1–15. Springer (2000)
6. Griffiths, G., Bittig, H.C., Johnson, K.S., Carval, T., Wong, A.P.S.: Ai4argo: Advancing automated quality control of argo ocean profiles using machine learning. Frontiers in Marine Science **9**, 862134 (2022). https://doi.org/10.3389/fmars.2022.862134
7. Idrees, M.M., Stahl, F., Badii, A.: Adaptive learning with extreme verification latency in non-stationary environments. IEEE Access **10**, 127345–127364 (2022). https://doi.org/10.1109/ACCESS.2022.3225225
8. Jayne, S.R., Roemmich, D., Zilberman, N., et al.: The argo program: Present and future. Oceanography **30**(2), 18–28 (2017). https://doi.org/10.5670/oceanog.2017.213
9. Johnson, K.S., Claustre, H., Takeshita, Y.: Observing biogeochemical cycles at global scales with profiling floats and gliders. Annual Review of Marine Science **13**, 23–43 (2021). https://doi.org/10.1146/annurev-marine-121219-081559
10. Kumm, M.M., Nolle, L., Stahl, F., Jemai, A., Zielinski, O.: On an artificial neural network approach for predicting photosynthetically active radiation in the water column. In: Bramer, M., Stahl, F. (eds.) Artificial Intelligence XXXIX. pp. 112–123. Springer International Publishing, Cham (2022)
11. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 413–422. IEEE (2008)
12. Lopez, C., Yang, F., Singh, R.: Machine learning methods for argo float qc: A comparative analysis. Journal of Atmospheric and Oceanic Technology **39**(9), 1451–1467 (2022)
13. López, A.M., Gonzalez, P., Shen, X.: Deep anomaly detection for argo profile data. In: IEEE OCEANS Conference (2022)
14. Roemmich, D., Gilson, J., Davis, R., et al.: The argo program: Observing the global ocean with profiling floats. Oceanography **22**(2), 34–43 (2009). https://doi.org/10.5670/oceanog.2009.36
15. Roemmich, D., Johnson, G.C., Riser, S., et al.: The global argo program: A decade of progress. Oceanography **36**(1), 22–33 (2023). https://doi.org/10.5670/oceanog.2023.110
16. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence **1**(5), 206–215 (2019)

17. Stahl, F.T., Nolle, L., Jemai, A., Zielinski, O.: A model for predicting the amount of photosynthetically available radiation from bgc-argo float observations in the water column. In: Proceedings of the ECMS. pp. 174–180 (2022)
18. Wong, A.P.S., Keeley, R., Carval, T., Team, A.D.M.: Argo quality control manual for ctd and trajectory data (2020). https://doi.org/10.13155/33951, version 3.4
19. Wong, A.P., Wijffels, S., Riser, S.C.: Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. Frontiers in Marine Science **7** (2020). https://doi.org/10.3389/fmars.2020.00700
20. Xie, J., Zhang, R., Wang, H.: Unsupervised anomaly detection in argo float data using isolation forest. Remote Sensing Letters **11**(4), 365–372 (2020)
21. Xie, J., Li, Z., Zhao, Z.: Detecting argo profile anomalies with autoencoders. Remote Sensing **12**(5), 817 (2020). https://doi.org/10.3390/rs12050817
22. Xie, W., Li, X., Du, J.: Anomaly detection for oceanographic profiles using autoencoder neural networks. IEEE Access **8**, 119660–119670 (2020)
23. Yigit, A., Klein, M., Douglass, E.: Interpretable ai for environmental monitoring: A review. Environmental Modelling Software **139**, 105031 (2021)
24. Zhang, K., Zhong, Z., Wang, Y.: Data standardization for deep learning: A survey. Neural Processing Letters (2019)