Article

Machine Learning for Enhanced Water Property Estimation in Coastal Environmental Monitoring

Gregory Murad Reis [1,†,‡], Luana Okino Sawada [1,‡], Paulo Victor Padrão [1,‡] and José Fuentes[1,‡]

[1] Knight Foundation School of Computing and Information Sciences, Florida International University, 11200 SW 8th St, Miami, FL 33199, Miami, FL, USA; gmuradre@fiu.edu
* Correspondence: gmuradre@fiu.edu; Tel.: +1(305)348-7852 (G.M.R.)
† Enviroment Finance & Risk Management, Florida International University, 11200 SW 8th St, Miami, FL 33199, Miami, FL, USA
‡ These authors contributed equally to this work.

Abstract: Coastal regions globally are experiencing rapid transformations due to climate change and anthropogenic impacts, necessitating robust monitoring systems. Autonomous surface and underwater robots (ASV and AUV) have emerged as important tools for high-resolution, large-scale data collection on water properties. However, these robots often face challenges from faulty or missing sensor data, affecting data accuracy and robot functionality. This paper explores the use of machine learning techniques to estimate water property parameters, addressing the challenges of missing or faulty data. By focusing on Biscayne Bay, Florida, this study uses linear regression, random forest, support vector regression, and multilayer perceptron, to predict parameters like dissolved oxygen, pH, and temperature. Initial results indicate the potential of these models to enhance data consistency and offer new perspectives for sensor fusion approaches.

Keywords: machine learning; water parameter estimation; coastal monitoring

Citation: Reis, G. M.; Okino Sawada, L.; Padrão, P. V.; Fuentes, J. Machine Learning for Enhanced Water Property Estimation in Coastal Environmental Monitoring. *Mathematics* **2023**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Introduction

In the face of climate change and anthropogenic impacts, coastal regions worldwide are undergoing rapid transformations, posing significant challenges to their fragile ecosystems [1]. These changes need robust and efficient monitoring systems to drive informed, timely decision-making for the preservation and management of these crucial habitats. Moreover, these changes may play a critical role in a country's infrastructure and surveillance [2]. Emerging technologies, such as autonomous surface and underwater robots (ASV and AUV), have become pivotal in the effort to monitor coastal waters, providing high-resolution, large-scale data on water property parameters, including total water column, temperature, pH, and dissolved oxygen levels [3–8].

However, these robotic systems often encounter challenges that may be caused by faulty or missing sensor data, significantly impacting the data collection process and potentially impairing the robot's functioning. To add to this, some of these sensing devices are expensive and sensitive, requiring regular maintenance to ensure accurate readings.

There have been several models proposed to address the issue of faulty sensor data, such as detection mechanisms in IoT systems [9,10]. Missing data estimation using machine learning is an important aspect of water quality research; however, the handling and reporting of missing data in prediction model studies using machine learning methods are often inadequate [11]. Many studies rely on deletion methods, such as complete-case analysis, which can introduce bias and reduce analytical power. Therefore, it is crucial for researchers to be aware of alternative methodologies to address missing data [11].

A personalized diagnosis method was proposed in [12] to detect faults in a bearing using acceleration sensors and finite element method (FEM) simulations. The method involved three steps and aimed to improve fault detection results. Machine learning techniques have also been used for fault diagnosis in other domains, such as in a pressurized water reactor (PWR) nuclear power plant [13]. Principal component analysis (PCA) models were employed to detect and diagnose sensor faults in the pressurizer of the PWR and an improved PCA-based method was proposed to successfully detect and isolate sensor faults, even in the presence of minor failures.

In [14], the authors evaluated the potential of remote sensing using machine learning techniques for improving water quality estimation over the coastal waters of Hong Kong. Concentrations of suspended solids (SS), chlorophyll-a (Chl-a), and turbidity were estimated with several machine learning techniques including Artificial Neural Network (ANN), Random Forest (RF), Cubist regression (CB), and Support Vector Regression (SVR). The results showed that machine learning algorithms can effectively estimate water quality parameters in inland lakes, with RF and SVR performing better than ANN.

Machine learning methods have shown promise in handling missing data and improving the estimation of water quality parameters. In [15], the authors proposed a novel water quality prediction model for a South African aquaculture farm using machine learning techniques that achieved accurate results with a low mean squared error. These methods have demonstrated outstanding imputation performance and provided methodological support for clinical decision-making in the presence of incomplete data [16]. The use of machine learning algorithms, such as RF, Support Vector Machine (SVM), and Artificial Neural Network (ANN), can effectively estimate water quality parameters in various settings, including inland lakes, rivers, and coastal waters.

Another study [17] focused on developing an efficient model using SVMs to predict the water quality (especially dissolved oxygen (DO) levels) of the Langat River Basin. Their proposed model analyzed data from six parameters of dual reservoirs in the catchment area and the SVM model was found to be effective in identifying the water quality status for the river catchment area. The study also discussed time-series predictive techniques for water quality, which utilize preceding time series and other parameters to predict water quality values. Various statistical analyses and AI-based modeling strategies have been used in these techniques for water quality prediction and water resources management. The models achieved high correlation coefficients and low prediction errors. Dual scenarios were employed to forecast water quality trends, with Scenario 1 validating the DO prediction scheme at each station and Scenario 2 validating the DO prediction scheme using information from prior stations. The models have also been found to be useful for those lacking sufficient monitoring stations for water quality parameters.

The application of machine learning techniques, combined with remote sensing data, has advanced the field of water environment monitoring. These methods have facilitated accurate water extraction and quantitative estimation of water quality and the integration of remote sensing big data, cloud computing, and machine learning has opened up new possibilities for monitoring and managing water resources [18].

In [19] the authors developed advanced artificial intelligence (AI) algorithms to predict water quality index (WQI) and water quality classification (WQC). The authors applied machine learning approaches like artificial neural networks (ANN), radial-basis-function (RBF), and regression to predict the chemical oxygen demand (COD) and water quality index (WQI).

Other machine learning methods, such as logistic regression, random forest (RF), Support Vector Machine, decision tree, k-nearest neighbor, XGBoost, gradient boosting, and naive Bayes, have been used in [20] for the continuous collection of water parameters data from sensors and the prediction of water quality. These machine learning models are implemented and tested to predict water quality attributes such as pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, turbidity, and potability.

This project aims to address these challenges by leveraging machine learning techniques to estimate water property parameters using available data from other parameters. By creating models capable of predicting these parameters, we can ensure consistent and reliable data collection even in the face of sensor failure or absence, thereby enhancing the autonomy of the marine robots.

In addition, by enabling more accurate and robust estimation of water property parameters, we aspire to facilitate better decision-making in the management of coastal ecosystems, promote the use of autonomous surface and underwater robots for environmental monitoring, and ultimately contribute to the sustainability of coastal regions and the preservation of their ecosystems. Particularly, this research focuses on Biscayne Bay, Florida, aiming to contribute to the ongoing efforts in environmental monitoring and protection of this vital coastal region.

The following report outlines the methodology, findings, and implications of our study. It details the data collection and preprocessing, the development and evaluation of various machine learning models, and the robustness of these models in dealing with missing or faulty sensor data. By sharing our research and findings, we hope to advance the field of marine robotics and contribute to a more sustainable future for our coastal ecosystems.

2. Materials and Methods

In this section, we outline the data collection and data processing procedures, as well as the step-by-step procedures employed to carry out the experiments.

2.1. Data Collection

Our study focused on a specific area within Biscayne Bay, specifically centered around the BBC campus of Florida International University (FIU). This region was chosen as a prime location of interest due to its proximity to FIU facilities. Moreover, it offers a convergence of factors for comprehensive research, including its close proximity to the urban area, active vessel traffic, and the presence of a thriving natural environment. Between September 2020 and November 2022, our research involved an extensive data collection initiative within the designated region of interest that resulted in more than 30 datasets in different seasons and months. To accomplish this, we employed the YSI Ecomapper [21], an autonomous underwater vehicle (AUV) that provides high-resolution water quality data, side scan sonar imaging, and bathymetric surveying. For every data collection mission, a variety of trajectories were planned, each designed to be completed within a timeframe ranging from 7 to 10 minutes. The AUV employed for data collection not only captured water parameters including temperature, pH, and dissolved oxygen, but also recorded vehicle parameters such as speed, longitude, latitude, and heading, especially important for troubleshooting and detailed analysis of missions. Data collected from each mission is then saved in CSV files for further processing.

Figure 1 shows the region of interest and trajectories performed by the autonomous data collection platform in different months in 2022, and Figure 2 shows the YSI Ecomapper AUV.

(a) **(b)**

Figure 1. Region of interest and data collection trajectories in (a) October 2022 and (b) November 2022.

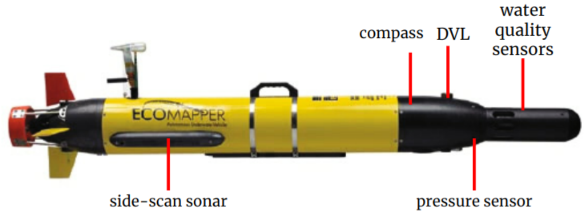

Figure 2. YSI Ecomapper, an autonomous underwater vehicle for high-resolution water quality data collection [21]

2.2. Data Processing 127

For each raw data sheet, we have organized the information into a dataframe consisting of 68 columns. These columns encompass a wide range of data, from internal AUV parameters and extending to various water parameters. To simplify and filter relevant information, the resulting dataframe primarily includes the AUV's position, indicated by longitude and latitude coordinates, timestamps (hh:mm:ss), and water parameters, including ODO (mg/L), Temperature (°C), pH, and Total Water Column (m). Table 1 shows a snippet of the resulting dataframe from November 2022 data set, Figures 3 and 4 provide the water parameter profiles, and Figure 5 shows the correlation between them. 128 129 130 131 132 133 134 135

Lat	Lon	Time	ODO(mg/L)	Temp.(°C)	pH	WaterColumn (m)
25.912771	-80.137886	12:00:32	4.21	28.400	8.27	0.60
25.912768	-80.137886	12:00:32	4.21	28.400	8.27	0.60
25.912767	-80.137883	12:00:33	4.20	28.400	8.27	0.61
25.912772	-80.137877	12:00:35	4.20	28.400	8.27	0.61
25.912766	-80.137879	12:00:35	4.20	28.400	8.27	0.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 1. Snippet of the dataset from November 2022 data collection mission.

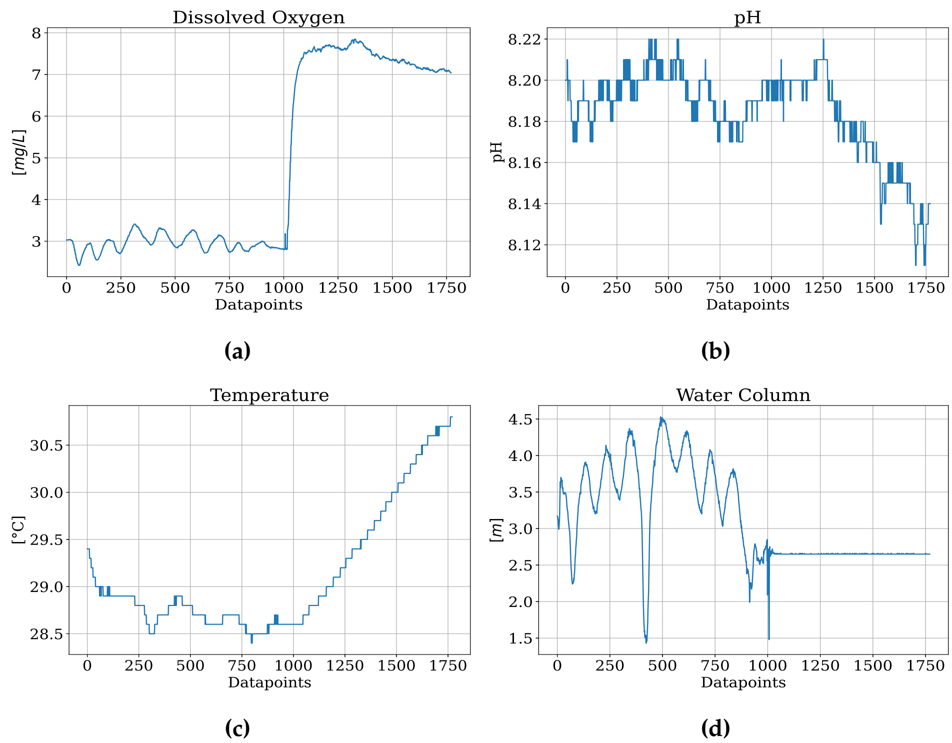

Figure 3. Water parameter profiles for the data collection mission in October 2022

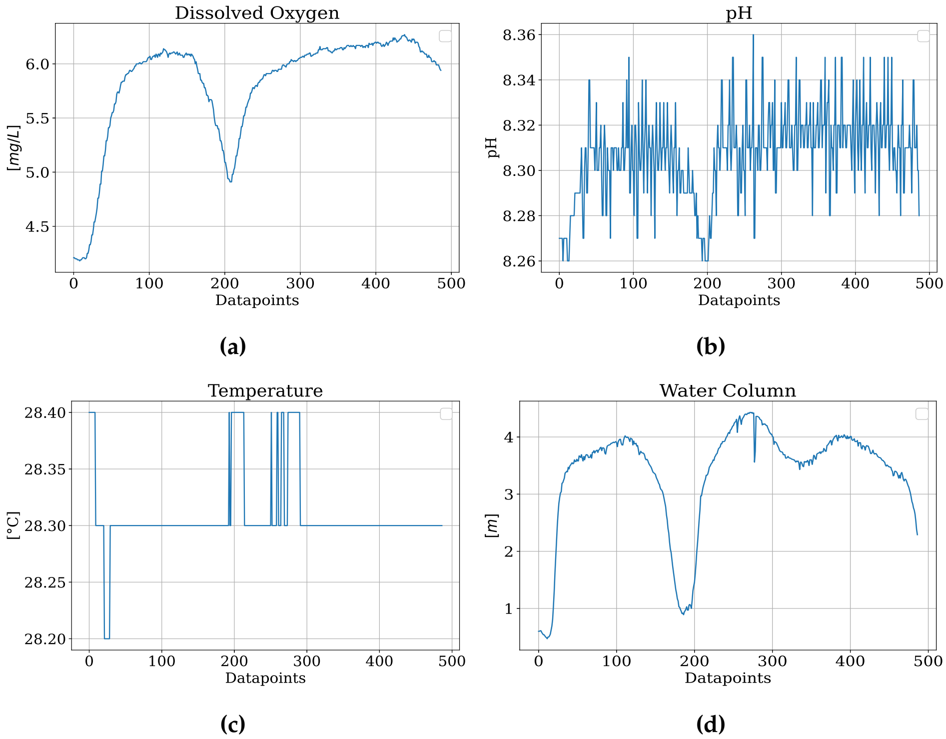

Figure 4. Water parameter profiles for the data collection mission in November 2022

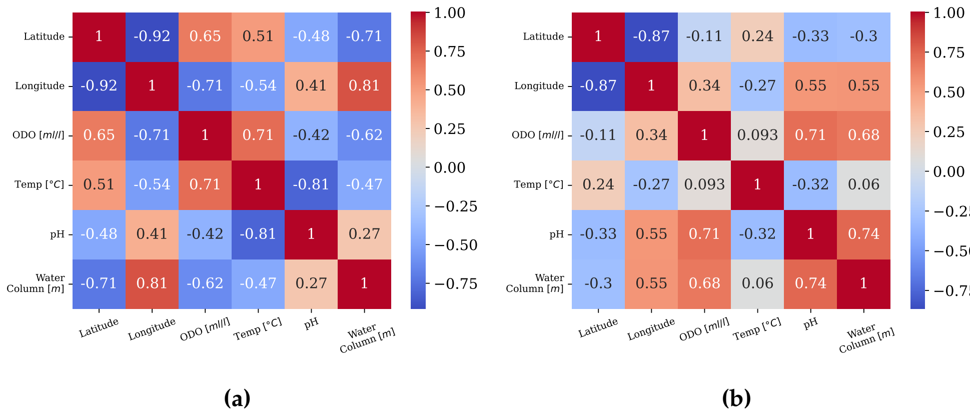

Figure 5. Correlation heatmaps for water parameters found at the data collection missions. 5a October 2022; 5b November 2022.

2.3. Machine Learning Methods

We tested different machine learning (ML) algorithms to compare various approaches and explore the relationship between water property parameters. We began with simple linear relationships, which were captured by linear regressors. As we progressed, we scaled up to more complex, non-linear relationships, captured by decision trees or non-linear kernels found in Support Vector Machines (in their regression version). Finally, we utilized Neural Networks as universal function approximators.

As mentioned earlier, we employed different approaches. Firstly, we used a linear regression (LR) model, which attempts to fit a line relating the features to the predicted value. Secondly, we utilized RFs, consisting of multiple decision trees that make predictions by considering individual features and assigning them thresholds. Thirdly, we employed Support Vector Machine (SVM) regressors (also known as SVR) with the kernel trick, providing a wide range of options for non-linear relationships. One key distinction from LR is that SVR allows for an error margin within which the predicted values are allowed to vary. Lastly, we employed Multi-layer Perceptron (MLP) architectures with varying numbers of neurons and layers to construct deep neural networks.

2.4. Simulation Methodology

In the previous section, we have introduced various methods. To characterize these methods comprehensively, we define a feature vector $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^5$, which includes both the water parameters and their corresponding longitude and latitude values. Furthermore, we introduce the target variable $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}$, representing one of the given parameters as shown in equation 1.

$$\begin{aligned}
(\mathbf{x}_1, \mathbf{y}_1) &= ((\text{lat, lon, temp, pH, water column}), \text{DO}) \\
(\mathbf{x}_2, \mathbf{y}_2) &= ((\text{lat, lon, DO, pH, water column}), \text{temp}) \\
(\mathbf{x}_3, \mathbf{y}_3) &= ((\text{lat, lon, temp, DO, water column}), \text{pH})
\end{aligned} \tag{1}$$

Then, we split the dataset into training and testing datasets, allocating 20% of the data for testing purposes. Additionally to the data preprocessing done before, we normalized the data by taking each feature x_i ($\mathbf{x}_j = (x_1, \ldots, x_5)$) and subtracting its mean and dividing it by its standard deviation in order to improve the ML training performance.

Hyperparameter tuning is known to be crucial for optimizing the performance of machine learning models, although it can be computationally expensive. Typically, it involves testing various parameter combinations to find the best ones (in the case of discrete parameters) and retraining the model with different hyperparameters (in the case of continuous parameters).

In our study, we selected a subset of hyperparameters to explore and performed an extensive grid search over each combination for each model. Linear regression is a simple model that does not have traditional hyperparameters. However, in some cases, the bias value can be fitted or fixed. We experimented with both options for this model.

For the RF models, we investigated different numbers of trees (n) in the forest, specifically $n \in 100, 200, 300$, and varied the maximum depth for each tree from no maximum depth to a maximum of 10 levels.

When working with SVR (Support Vector Regression) models, we employed Ridge parameter regularization based on the L_2 norm. The regularization constant was selected from the set $0.1, 1, 10$ to prevent overfitting. Additionally, we experimented with different kernel functions, including linear and radial-basis functions.

Regarding the MLP (Multilayer Perceptron) model, we explored different fully connected architectures, considering 1, 2, and 3 layers with 100 neurons in each layer. Furthermore, we tested various activation functions such as tanh, sigmoid, and relu. We also experimented with different optimizers and learning rates, specifically 0.0001, 0.001, and 0.01.

3. Results and Discussion

In this section, we present empirical findings emanating from the application of four distinct machine-learning paradigms to two discrete datasets, collected in October 2022 and November 2022, respectively. The outcomes of our computational experiments are succinctly encapsulated within Figure 6 and Figure 7, providing a graphical overview of predictive outcomes. To quantitatively assess our proposed models' efficacy, we harnessed quintessential performance metrics, encompassing Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE), Mean Absolute Error (MAE), and R-squared. These well-established metrics, esteemed within the realm of Data Science, offer pivotal insights into the predictive precision and prowess of our models, thereby establishing a foundation for well-informed data-driven decisions. The ensuing tabular presentations meticulously encapsulate quantified outcomes, spotlighting each model's performance concerning specific water parameters.

Within the context of these findings, it becomes evident that predictive modeling of water features is indeed feasible through strategic exploitation of interrelated feature dynamics, contingent upon the judicious selection of the appropriate ML algorithm. Exemplifying this point, Support Vector Regression (SVR) exhibits comparable performance to other algorithms in predicting dissolved oxygen levels when considering temperature, pH, and sample collection depth. However, the investigation reveals SVR's limitation in estimating temperature and pH using their corresponding counterparts. This limitation might stem from SVR's design, which permits an error margin allowing for mispredictions. Conversely, the absence of such a margin in LR could lead to the assignment of constant values within it. In contrast, the RF algorithm consistently demonstrates reliable results across all estimations, reflecting its capacity to capture latent relationships and perform well for each water feature.

Figure 6 and Figure 7 synthesize the predictive proficiency of distinct algorithms against actual values within the testing datasets. These figures align each algorithm with a specific row and each column with an individual water parameter.

Further insights emerge from Tables 2 through 7, which detail metric evaluations for dissolved oxygen, pH, and temperature estimations. Focusing on dissolved oxygen estimation for both October 2022 (Table 2) and November 2022 (Table 5), RF and MLP consistently excel across proposed metrics. Their precision, exemplified by lower MSE, RMSE, NRMSE, MAE, and elevated R-squared values, underscores their competence in capturing dissolved oxygen dynamics.

Analogously, in pH estimation, spanning October (Table 3) and November 2022 (Table 6), RF and MLP consistently demonstrate a superior grasp of underlying data patterns. This efficacy is highlighted through their ability to maintain comparably low MSE, RMSE, NRMSE, MAE, and high R-squared values, affirming their aptitude for decoding pH fluctuations.

Regarding temperature estimation across October (Table 4) and November 2022 (Table 7), the models – particularly RF and MLP – consistently exhibit heightened precision. Their efficacy, as evidenced by their relatively lower MSE, RMSE, NRMSE, MAE, and elevated R-squared values, solidifies their role in decoding temperature dynamics.

In conclusion, initial experiments have yielded promising results, unveiling new avenues for sensor estimation and calibration techniques. These findings offer the potential to estimate data by leveraging various water features, providing a valuable contrast to measurements derived from raw data. This expansion of perspectives enriches the scope of sensor fusion approaches, e.g. Extended Kalman Filters, as the outcomes indicate that water features are not mutually orthogonal; rather, they exhibit interconnected relationships that enhance the consistency of data collection and processing missions. Also, the data analysis expands the understanding of the predictive potential inherent within diverse machine-learning models for estimating water parameters. These insights significantly contribute to unraveling the multifaceted implications of water parameter oscillations on coastal ecosystems.

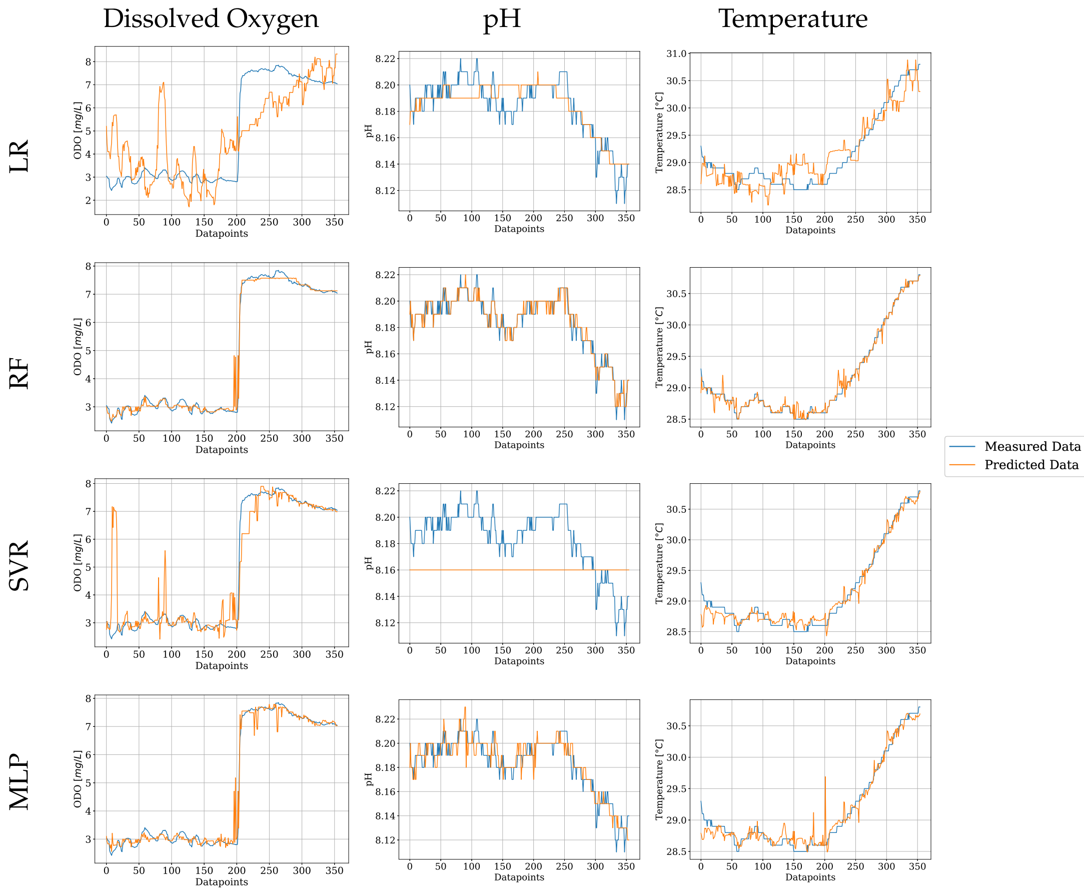

Figure 6. Summarized results contrasting the predicted and real values on the testing set, each row considers one ML algorithm; each column is a water feature. Data were collected in October 2022, see Figure 1a.

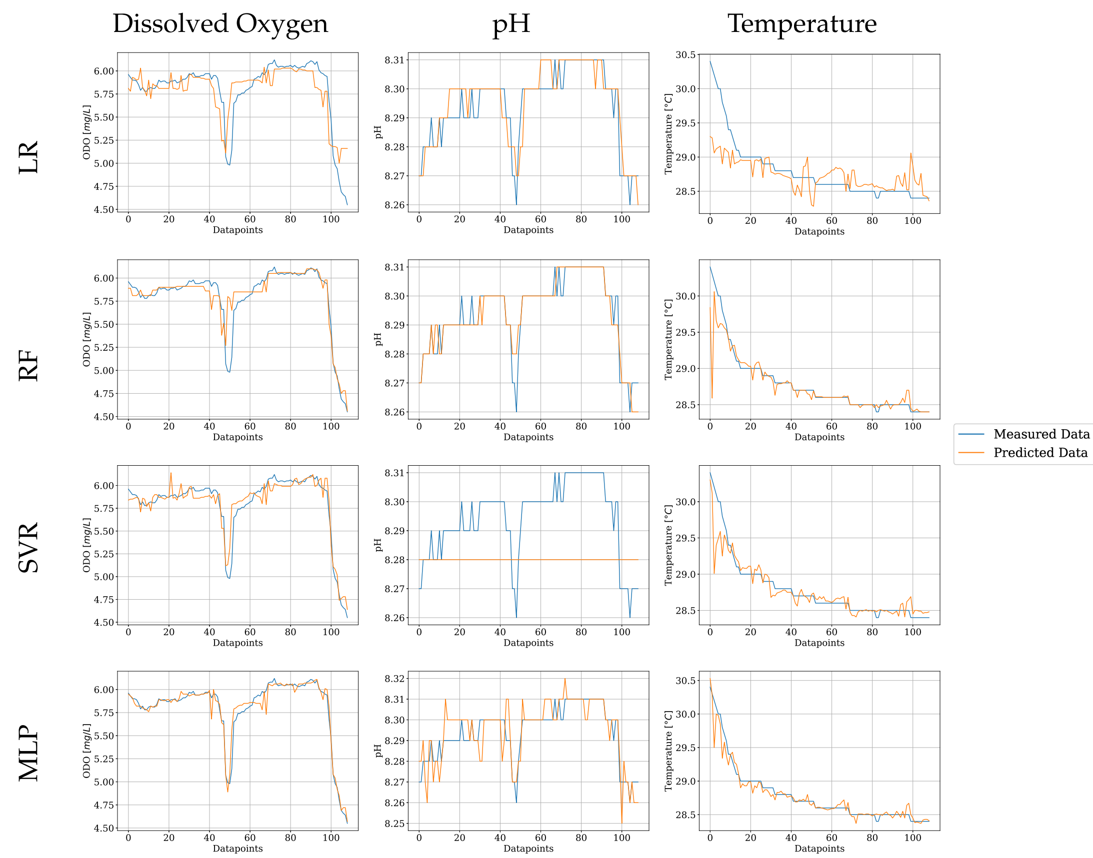

Figure 7. Summarized results contrasting the predicted and real values on the testing set, each row considers one ML algorithm; each column is a water feature. Data were collected in November 2022, see Figure 1b.

Dissolved Oxygen	LR	RF	SVR	MLP
MSE	1.921	**0.044**	0.616	0.065
RMSE	1.3860	**0.210**	0.785	0.256
NRMSE	25.525	**3.884**	14.463	4.716
MAE	1.101	**0.105**	0.341	0.132
R-squared	0.608	**0.991**	0.874	0.987

Table 2. Evaluation metrics for dissolved oxygen estimation for the data collected in October 2022, see Figure 1a.

pH	LR	RF	SVR	MLP
MSE	0.0001	**0.00002**	0.0009	0.0001
RMSE	0.011	**0.005**	0.029	0.009
NRMSE	10.844	**4.819**	26.884	8.165
MAE	0.009	**0.003**	0.027	0.007
R-squared	0.727	**0.946**	-0.679	0.845

Table 3. Evaluation metrics for pH estimation for the data collected in October 2022, see Figure 1a.

Temperature	LR	RF	SVR	MLP
MSE	0.076	**0.006**	0.014	0.016
RMSE	0.276	**0.078**	0.116	0.126
NRMSE	12.008	**3.384**	5.055	5.461
MAE	0.232	**0.042**	0.088	0.091
R-squared	0.838	**0.987**	0.971	0.966

Table 4. Evaluation metrics for temperature estimation for the data collected in October 2022, see Figure 1a.

Dissolved Oxygen	LR	RF	SVR	MLP
MSE	0.041	0.20	0.011	**0.005**
RMSE	0.203	0.141	0.109	**0.074**
NRMSE	12.944	9.041	6.968	**4.713**
MAE	0.136	0.071	0.077	**0.043**
R-squared	0.666	0.837	0.903	**0.955**

Table 5. Evaluation metrics for dissolved oxygen estimation for the data collected in November 2022, see Figure 1b.

pH	LR	RF	SVR	MLP
MSE	0.00003	**0.00001**	0.00036	0.00005
RMSE	0.005	**0.004**	0.019	0.007
NRMSE	11.298	**8.302**	38.408	15.352
MAE	0.004	**0.002**	0.017	0.005
R-squared	0.816	**0.901**	-1.115	0.662

Table 6. Evaluation metrics for pH estimation for the data collected in November 2022, see Figure 1b.

Temperature	LR	RF	SVR	MLP
MSE	0.099	0.037	0.031	**0.010**
RMSE	0.315	0.194	0.176	**0.104**
NRMSE	15.779	9.740	8.815	**5.210**
MAE	0.204	0.065	0.095	**0.055**
R-squared	0.478	0.801	0.837	**0.943**

Table 7. Evaluation metrics for temperature estimation for the data collected in November 2022, see Figure 1b.

4. Conclusions

Our study explores the transformative potential of machine learning techniques in enhancing coastal water monitoring, especially in the face of sensor data challenges. By applying models such as LR, RF, and SVR, we achieved promising results in predicting critical water parameters like dissolved oxygen, pH, and temperature. Particularly in Biscayne Bay, Florida, these methodologies not only ensure data consistency but also pave the way for more advanced sensor fusion approaches. As coastal regions continue to face environmental challenges, such innovative solutions will be important in ensuring robust and informed monitoring for the preservation of these vital ecosystems. Future work should aim to augment the number of water features to be entangled and estimated by this approach. Moreover, it should state and corroborate how necessary seasonal models are or if on the other hand, spatio-temporal models are required to keep a meaningful estimation giving rise to more accurate and interpretable models [5,22].

Author Contributions: Conceptualization, Gregory Murad Reis; methodology, Gregory Murad Reis and Luana Okino Sawada; software, Gregory Murad Reis and Luana Okino Sawada; validation, Paulo Victor Padrão and José Fuentes; formal analysis, Paulo Victor Padrão and José Fuentes; investigation, Gregory Murad Reis and Luana Okino Sawada; resources, Gregory Murad Reis; data curation,

Gregory Murad Reis and Paulo Victor Padrão; writing—original draft preparation, Gregory Murad Reis, Paulo Victor Padrão and José Fuentes; writing—Paulo Victor Padrão, José Fuentes and Luana Okino Sawada; visualization, José Fuentes; supervision, Gregory Murad Reis; project administration, Gregory Murad Reis; funding acquisition, Gregory Murad Reis All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by funding from the Institute of Environment Environmental Finance and Risk Management Program at Florida International University.

Data Availability Statement: In alignment with the recommendations of MDPI journals, we are committed to promoting transparency and openness in research. We are pleased to make the data supporting the reported results of this study available to the broader scientific community. Interested researchers can access the datasets analyzed or generated during our study at our dedicated platforms: www.oceanrobotics.us and https://fiumarinerobotics.azurewebsites.net/. We believe that sharing our data will foster collaboration, enhance reproducibility, and further the advancements in the field.

Acknowledgments: We would like to acknowledge Dr. Leonardo Bobadilla and Dr. Ryan N. Smith for making the robots available and for their technical insights and thoughts.

References

1. Miloslavich, P.; Bax, N.J.; Simmons, S.E.; Klein, E.; Appeltans, W.; Aburto-Oropeza, O.; Andersen Garcia, M.; Batten, S.D.; Benedetti-Cecchi, L.; Checkley Jr, D.M.; et al. Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. *Global Change Biology* **2018**, *24*, 2416–2433.
2. Alam, T.; Campaneria, A.; Silva, M.; Bobadilla, L.; Weaver, G.A. Coastal Infrastructure Monitoring through Heterogeneous Autonomous Vehicles. In Proceedings of the 2020 Fourth IEEE International Conference on Robotic Computing (IRC), 2020, pp. 79–82. https://doi.org/10.1109/IRC.2020.00019.
3. Aguzzi, J.; Albiez, J.; Flögel, S.; Godø, O.R.; Grimsbø, E.; Marini, S.; Pfannkuche, O.; Rodriguez, E.; Thomsen, L.; Torkelsen, T.; et al. A Flexible Autonomous Robotic Observatory Infrastructure for Bentho-Pelagic Monitoring. *Sensors* **2020**, *20*. https://doi.org/10.3390/s20061614.
4. Rojas, C.A.; Reis, G.M.; Albayrak, A.R.; Osmanoglu, B.; Bobadilla, L.; Smith, R.N. Combining Remote and In-situ Sensing for Persistent Monitoring of Water Quality. In Proceedings of the OCEANS 2022-Chennai. IEEE, 2022, pp. 1–6.
5. Rojas, C.A.; Padrao, P.V.; Fuentes, J.E.; Albayrak, A.R.; Osmanoglu, B.; Bobadilla, L. Combining Remote and In-situ Sensing for Autonomous Underwater Vehicle Localization and Navigation. In Proceedings of the OCEANS 2022, Hampton Roads, 2022, pp. 1–7. https://doi.org/10.1109/OCEANS47191.2022.9977208.
6. Suwadi, N.A.; Derbali, M.; Sani, N.F.M.; Lam, M.C.; Arshad, H.; Khan, I.; Kim, K. An Optimized Approach For Predicting Water Quality Features Based On Machine Learning. *Wireless Communications and Mobile Computing* **2022**, *2022*, 1–20. https://doi.org/10.1155/2022/3397972.
7. Torky, M.; Bakhiet, A.; Bakrey, M.; Ismail, A.A.; Seddawy, A.I.B.E. Recognizing Safe Drinking Water and Predicting Water Quality Index Using Machine Learning Framework. *International Journal of Advanced Computer Science and Applications* **2023**, *14*. https://doi.org/10.14569/ijacsa.2023.0140103.
8. Padrao, P.; Fuentes, J.; Bobadilla, L.; Smith, R.N. Estimating spatio-temporal fields through reinforcement learning. *Frontiers in Robotics and AI* **2022**, *9*. https://doi.org/10.3389/frobt.2022.878246.
9. Malik, A.D.; Jamil, A.; Omar, K.A.; Abd, M.H. Implementation of faulty sensor detection mechanism using data correlation of multivariate sensor readings in smart agriculture. *Annals of Emerging Technologies in Computing (AETiC)* **2021**, *5*.
10. Gaddam, A.; Wilkin, T.; Angelova, M.; Gaddam, J. Detecting sensor faults, anomalies and outliers in the internet of things: A survey on the challenges and solutions. *Electronics* **2020**, *9*, 511.
11. Nijman, S.; Leeuwenberg, A.; Beekers, I.; Verkouter, I.; Jacobs, J.; Bots, M.L.; Asselbergs, F.W.; Moons, K.G.M.; Debray, T. Missing Data Is Poorly Handled and Reported In Prediction Model Studies Using Machine Learning: A Literature Review. *Journal of Clinical Epidemiology* **2022**, *142*, 218–229. https://doi.org/10.1016/j.jclinepi.2021.11.023.

12. Liu, X.; Huang, H.; Xiang, J. A Personalized Diagnosis Method To Detect Faults In a Bearing Based On Acceleration Sensors And An Fem Simulation Driving Support Vector Machine. *Sensors* **2020**, *20*, 420. https://doi.org/10.3390/s20020420.
13. Naimi, A.; Deng, J.; Doney, P.; Akbari, A.S.; Shimjith, S.R.; Arul, A.J. Machine Learning-based Fault Diagnosis For a Pwr Nuclear Power Plant. *IEEE Access* **2022**, *10*, 126001–126010. https://doi.org/10.1109/access.2022.3225966.
14. Hafeez, S.; Wong, M.; Ho, H.C.; Nichol, J.E.; Nichol, J.E.; Abbas, S.; Tang, D.; Lee, K.S.; Pun, L. Comparison Of Machine Learning Algorithms For Retrieval Of Water Quality Indicators In Case-ii Waters: a Case Study Of Hong Kong. *Remote Sensing* **2019**, *11*, 617. https://doi.org/10.3390/rs11060617.
15. Eze, E.C.; Halse, S.; Ajmal, T. Developing a Novel Water Quality Prediction Model For A South African Aquaculture Farm. *Water* **2021**, *13*, 1782. https://doi.org/10.3390/w13131782.
16. Wang, H.; Tang, J.; Wu, M.; Wang, X.; Zhang, T. Application Of Machine Learning Missing Data Imputation Techniques In Clinical Decision Making: Taking the Discharge Assessment Of Patients With Spontaneous Supratentorial Intracerebral Hemorrhage As An Example. *BMC Medical Informatics and Decision Making* **2022**, *22*. https://doi.org/10.1186/s12911-022-01752-6.
17. Yahya, A.S.A.; Ahmed, A.; Othman, F.; Ibrahim, R.K.; Afan, H.A.; El-Shafie, A.H.; Chow, M.F.; Hossain, S.; Ehteram, M.; El-Shafie, A. Water Quality Prediction Model Based Support Vector Machine Model For Ungauged River Catchment Under Dual Scenarios. *Water* **2019**, *11*, 1231. https://doi.org/10.3390/w11061231.
18. Jinyue, C.; Fu, R.; Chen, S.; Li, D.; Jiang, H.; Wang, C.; Peng, Y.; Jia, K. Remote Sensing Big Data For Water Environment Monitoring: Current Status, Challenges, and Future Prospects. *Earth's Future* **2022**, *10*. https://doi.org/10.1029/2021ef002289.
19. Aldhyani, T.H.H.; Al-Yaari, M.; Alkahtani, H.; Maashi, M.S. Water Quality Prediction Using Artificial Intelligence Algorithms. *Applied Bionics and Biomechanics* **2020**, *2020*, 1–12. https://doi.org/10.1155/2020/6659314.
20. Sen, S.; Maiti, S.; Manna, S.; Roy, B.; GHOSH, A. Smart Prediction Of Water Quality System For Aquaculture Using Machine Learning Algorithms **2023**. https://doi.org/10.36227/techrxiv.22300435.
21. YSI EcoMapper Autonomous Underwater Vehicle, https://www.ysi.com/ecomapper, 2023.
22. Fuentes, J.; Bobadilla, L.; Smith, R.N. Localization in Seemingly Sensory-Denied Environments through Spatio-Temporal Varying Fields. In Proceedings of the 2022 Sixth IEEE International Conference on Robotic Computing (IRC). IEEE, 2022, pp. 142–147.