A Water Quality Research Platform for the Near-real-time Buoy Sensor Data

Maria Presa-Reyes*, Biayna Bogosian†, Bradley Schonhoff‡, Christopher Jerauld*,
Christian Moreyra*, Piero Gardinali§, Shu-Ching Chen*
*School of Computing and Information Sciences
Florida International University, Miami, Florida 33199
Emails: {mpres029,cjera004,cmore100,chens}@cs.fiu.edu
†Department of Architecture
Florida International University, Miami, Florida 33199
Email: bbogosia@fiu.edu
‡Institute of Environment
Florida International University, Miami, Florida 33199
Email: bschonho@fiu.edu
§Department of Chemistry and Biochemistry
Florida International University, Miami, Florida 33199
Email: gardinal@fiu.edu

***Abstract*—Maintaining environmental sustainability relies on continuously monitoring environmental conditions. Water is an environmental component essential to the survival of all living organisms; hence, to prevent contamination and ensure proper water treatment, persistent observations and measurements of water quality are crucial. Traditionally, the procedure for testing the quality of water involved traveling to designated testing sites, manually collecting surface samples, transporting said samples to a laboratory for analysis, analyzing chemicals and microbial contaminants, and publishing the findings with the community. The technological advances in wireless sensor networks bring forth the opportunity for remote measurement and monitoring of water samples. Not only is the presence of the scientist no longer mandatory on the testing site, but the data can also be automatically collected, visualized, monitored, and shared through sensor recordings. These transitions exhibit a much fine-grained level of spatio-temporal information collection and allow for more comprehensive and long-term studies. Three research buoys, designed to be deployed in both shallow freshwater ecosystems and near-shore marine environments, were launched in different locations of South Florida to tackle complex challenges of environmental contamination. The research presented here designs and deploys a water quality monitoring platform for allowing the scientists to analyze better the near-real-time data collected by the buoys and generate insights. We further demonstrate two engaging near-real-time visualization methods developed to disseminate data trends and findings to a wide range of audiences from diverse backgrounds.**

***Keywords*-water quality monitoring; data management; sensor buoys; augmented reality; data visualization;**

I. Introduction

It is well-documented that the increase of the human population has had a negative impact on the environment [1]. One of the most widespread, costly, and challenging environmental problems in the United States is nutrient pollution [2], which is often caused by the excess of Nitrogen and Phosphorus as a result of a wide range of human activities. Nitrogen and Phosphorous are essential parts of the ecosystem, and both support the growth of aquatic plants, which provide food and habitat to marine animals. Meanwhile, Nitrogen is the most abundant nutrient in the air we breathe. However, too much of these nutrients cause an imbalance in the environment, whereby air and water can become polluted. The uncontrollable growth of algae in the ecosystem as a result of the excess nutrients creates harmful repercussions to water quality as well as the aquatic life.

Monitoring measurements of nutrients such as Nitrogen and Phosphorous in the ecosystem helps prevent contamination, ensures proper water treatments, and maintains environmental sustainability. Environmental monitoring and research are achieved through continuous sampling and analysis of changes in environmental conditions. There is a variety of sampling methods, and some of the most common environmental samples include air, water, soil, biological materials, and wastes [3]. Choosing a sampling method and the frequency when sampling is done highly depends on the study's objectives. Water quality monitoring typically involves periodic sampling and analyzing water conditions about its chemical, physical, and biological properties. The traditional approach to monitor water quality involved manual effort, where an individual required to make constant trips to designated testing sites and transport *grab samples* to a laboratory for analysis and detection of chemicals and contaminants [4].

Instrumentation crafted to monitor the environmental conditions continuously became an ideal approach for some environmental measurements [3]. The technological advances of wireless sensor networks bring forth the opportunity for remote monitoring and analysis of most of these samples. Water data is automatically collected and monitored through sensor recordings exhibiting a much fine-grained level of spatio-temporal information and allowing for more comprehensive and long-term studies.

Literature has shown tremendous advancements in the procedure for sampling and collection of the data through wireless sensor networks [5], [4], [6]. However, few studies tackle the technologies and methods that can be utilized to make data acquisition more efficient while assisting the scientists in the analyzes and communication of the findings. Once data has been collected, it is often up to the scientist to perform pre-processing, data exploration, and integration of data from different sources to apply the analytics that will lead to insights and conclusions. Our proposed platform goes further than data storage and retrieval and brings more advanced techniques in data integration, monitoring, and the applications that have resulted from such a platform. The objective is to develop a robust platform that will support an emerging and scalable network for near-real-time wireless water quality monitoring with water buoys.

The main contributions of this paper are summarized as follows:

- A novel application of water quality monitoring through the integration of different sensors into a relatively-compact and portable buoy unit is introduced.
- A novel research platform is designed to address both the specific challenges concerning the buoy-collected data as well as the more general challenges pertaining sensor-based water quality monitoring that have yet to be tackled by the literature.
- Two novel applications are built on top of the proposed platform to facilitate data exploration, and information visualization, while aiming to reach a wide range of audiences from diverse backgrounds.

The remainder of this paper starts with section II demonstrating the specifications and motivation behind the deployment of the water buoys. Section III reviews the recent development in water quality monitoring sensor devices along with relevant platforms designed to manage the sampled data. Section IV introduces the technical details of our proposed platform. Section V describes the different applications resulting from the proposed platform. Section VI proposes some future directions for further advancement in the water quality research using buoy sensor data. Finally, Section VII concludes the paper.

II. Research Buoys

In order to address water contamination in South Florida, the NSF (National Science Foundation) CREST (Center of Research Excellence in Science and Technology) Center for Aquatic Chemistry and Environment (CAChE) at Florida International University has deployed three cutting-edge research buoys, specially-designed to be utilized in both shallow freshwater ecosystems and near-shore marine environments. Developed by Xylem Inc., our customized system features the EMM150 Research Buoy platform equipped with YSI's EXO2 water quality sonde and a Doppler Current Sensor by Aanderaa. The buoys measure the following parameters: temperature, conductivity (salinity), dissolved oxygen, pH, turbidity, chlorophyll, depth, and directional water flow rates. All sensors are maintained automatically by a self-cleaning wiper system that prevents biofouling and allows for long-term deployment. Additionally, the buoys are equipped with a set of solar panels, which recharge the central battery and can extend sampling periods to several months. Data collected are automatically transmitted via cellular uplink every 15 minutes, making the information available in nearly real-time to our CREST CAChE team and the general public.

The integration of these sensors into a relatively-compact buoy platform represents a novel application for water quality monitoring; the buoys are able to be deployed in a range of aquatic ecosystems of varying depths, and our researchers can receive near-real-time indicators of potential contaminants and other environmental issues remotely. Furthermore, the water-current data can provide useful information for tracking contaminants' pathways, indicating the general direction of the source and its trajectory. These buoys are placed in diverse locations around southeast Florida to monitor water quality and collect baseline data in our nearshore ecosystems and urban waterways. Our first buoy deployment in October 2018 was in response to a red tide outbreak on the southeast coast of Florida. Since then, we have maintained a buoy at this site in Haulover Inlet, located in the northern area of Biscayne Bay, to continue monitoring this unique environment, which receives a mix of freshwater inputs and tidal exchanges. Next, we deployed a buoy in a Coral Gables canal to assess and compare conditions in an urban environment. Our third buoy was temporarily deployed within a harbor in the Florida Keys—where harmful algae and fish kills had been reported—and then a mangrove forest in Biscayne Bay. This buoy was later deployed for long-term monitoring in the Miami River, arguably one of the most polluted waterways in south Florida. Each of these locations represents a unique environment, and the use of these research buoys provides valuable information on the conditions within our aquatic ecosystems, which our team can then share with public audiences and decision-makers.

III. Related Work

The increase in pollution and water contamination has prompted the deployment of water sensor networks to monitor a continuous measurement of water quality [7]. Different types of environmental sensors have been developed and are continued to be developed and deployed throughout the years. Some examples range from the traditional and commercial sensors, such as the SEBA multi-parameter sensors demonstrated in [5], to the more low-cost, low-power, and

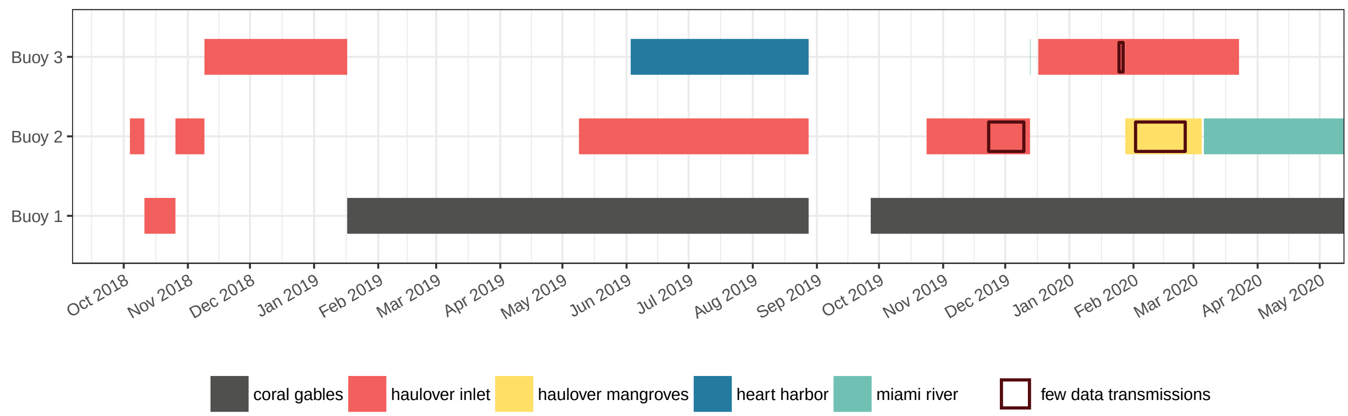

Figure 1. Timeline of buoy activity.

do-it-yourself (DIY) environmental sensing and data logging sensor developed using open-source hardware [6].

In addition, various online platforms have been proposed to automatically analyze the water quality data collected by either sensors or other sampling methods, and detect potential issues. Horsburgh *et al.* in 2019 [6] proposed a web-based data sharing and visualization portal to support a network of Arduino-based water quality data loggers. Previously proposed web-based meta-databases have also aimed to reconcile the heterogeneity in data from different sources to use for analytics; however, the integration is mostly superficial where it will not properly handle the integration between discrete and continuous spatio-temporal data by applying the appropriate transformations [8], [9].

Previously introduced studies demonstrate the rapid developments in the water quality data collection through various sensors as well as different approaches for handling the data collected, which usually focuses on storage, retrieval, and some straightforward visualizations. However, integration of the data collected considering the diverse factors reflected in the data (i.e., time, sampling site, sampling frequency, water source) remains an open issue [4]. The proposed platform was developed to support the advancements in water quality monitoring and research from parameters measured by the nearly live feed of three water buoys equipped with multi-parameter sensors. Our objective is to propose new algorithms and methods that serve to overcome the difficulties that arise from working with the sensor-based data collected.

IV. Proposed Platform

The core of the proposed platform (shown in Figure 2) is to make the data more accessible and easy to manage to support advancements in water quality monitoring research from the data acquired by an emerging network of buoy sensors. Data collected by the buoys is foremost signaled to HydroSphere, a cloud-based platform developed by Xylem

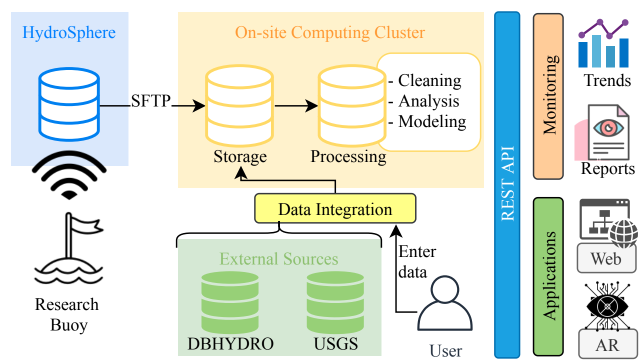

Figure 2. Proposed platform.

to store and visualize the measurements sensed from the water. As new data becomes available, it is transferred and backed-up to a server on-site through SFTP (Secure File Transfer Protocol), to allow for further analysis and research. The platform follows an extract, load, transform (ELT) approach—raw data is first transferred to the on-site computing cluster, and transformations are then applied to prepare the data for sharing, integration, and analysis. A REST API allows for the processed data accessible to the users, who can then retrieve the relevant information to conduct studies and/or build applications based on the buoy data.

A. Data Management

1) Data Storage: The current database (Figure 3) is designed to store water quality measurements and their supporting metadata in a structured way while considering scalability. It is capable of efficiently guaranteeing data integrity and avoiding redundancy through a normalized design. Applications built on top of the database can leverage its design to query, analyze, and manage the data efficiently and effectively. The extensive normalization process involved in the design of the database, allows developers and administrators to effectively scale and accommodate new entities, data, and its relationships.

The database was implemented in PostgreSQL. The design of the data model seeks to set a standard across the platform that will allow the integration of the buoy data with data from various sources. Tables dedicated to defining parameter measurements, sampling methods, and unit of measure, serve to support a consistent terminology. New terminologies can be added as long as they are approved by an administrator, who will ensure there is no duplicates in the database and all the information submitted is correct.

One of the buoy's defining features is their ability to be deployed and transferred to different testing sites. A summary of the buoy activity timeline is demonstrated in Figure 1. The data model takes into consideration the variability of the locations of each buoy unit by keeping records of the buoy's location changes and the respective timestamp. The user does not need to worry about learning the buoy activity timeline when requesting all data sampled by the buoys at a specific location.

2) Web Service: Our platform offers a RESTful Application Programming Interface (API) that makes the data that has been processed more accessible and suitable for other applications through the web. Registered users can request an API key to access the data and models available through our Representational State Transfer (REST) API. An API is a computing interface that defines a set of standard of interactions between different software to promote more interoperable communications and data sharing [10].

The data requested is delivered in the universal JavaScript Object Notation (JSON) format [11]. JSON is supported by most browsers and can be read and parsed by all programming languages considering JSON is text only. Users with access to our API can retrieve data over the web and define data parameters of interest in instances where very specific data is required. Moreover, various applications, such as data exploration, evaluation, monitoring, can leverage the most up-to-date data that has already been cleaned, validated, and processed.

The REST API offers various functionalities through the GET request, including but not limited to (1) retrieval of the newest measurements a buoy at a defined location has collected, (2) retrieval of all the data collected from a specific day by from a defined location, (3) retrieval of the data collected from a certain day and time at a defined location, and (4) retrieval of some specific measurements sampled by the a buoy at a defined location.

B. Data Processing

Data transformation, cleaning, and aggregation are some of the processing techniques required to properly prepare data for analysis. Methods offered by the platform include (1) *Data cleaning*: Handle cases such as missing data and corrupted values; (2) *Format revision*: Detecting inconsistency and ensure data values abide by the rules for its field type defined by the metadata; (3) *Standardization*: Validate unit of measurements, date and time and bring data into a standard format that facilitates integration and collaborative research.

On some occasions, often depending on the challenges introduced by the sampling environment, specific issues arise that may lead to inaccurate values in the sensed readings that would require the intervention of the domain expert to validate and flag these misleading data points. Some well-known issues when it comes to sensor-based water quality monitoring include bio-fouling, sensor drift, and underwater signal propagation [4]. However, working with near-real-time data places some overhead in the data cleaning process. A year-worth of information collected from a single location signifies thousands of data points that require review and validation. The platform automates many of the data transformation techniques to expedite the data cleaning process and reduce the expert's workload. A preliminary label is assigned to the data points that need further reviewing.

A sudden spike or dip in a time-series data is anomalous behavior. However, not all anomalies can be classified as noise or mistakes in the data readings. The platform applies the Isolation Forest algorithm [12], [13] to detect anomalies in the data following an unsupervised approach. Isolation Forest *isolates* each data instance by randomly singling out a feature and then randomly selecting a split value between the maximum and minimum values of the chosen feature. Each point returns an anomaly score, and those points with a higher score are then reviewed by a domain expert who will decide whether the point should be removed before further analysis is made.

C. Data Integration

Integration is a crucial step that serves to reconcile data from several sources and provide a unified view. To solve, or at least mitigate, the data-interoperability problem, some principles and expectations about the data must first be defined. All the data uploaded and processed in the environment must be accompanied with the appropriate metadata such that another group/researcher is able to reproduce the published results [14].

As stated in [4], the integration of water quality parameters from different sampling points remains an open issue. Our platform is equipped with the methodologies necessary to prepare the data for integration and analysis. A user can make decisions on which data he/she wishes to integrate and what type of transformations the data must go through before integration, including techniques such as subsampling, aggregation, interpolation, and conversion of units of measurements when necessary. The environment is highly heterogeneous; hence, spatial, and temporal variability is

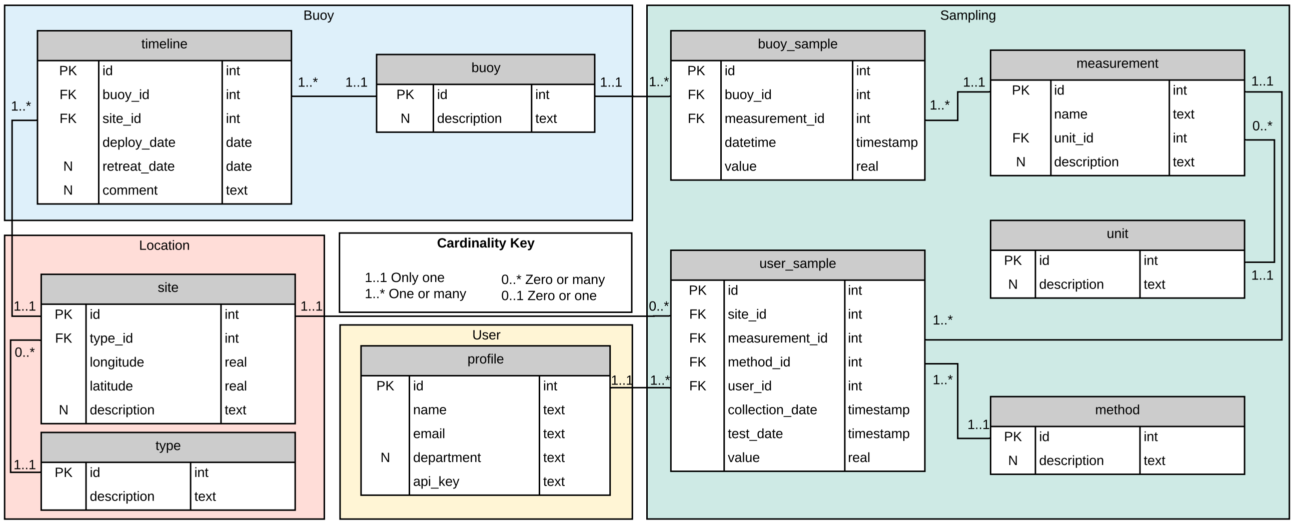

Figure 3. Logical data model. The **PK** label depicts the primary key field for each table. The **FK** label depicts each table's foreign key(s). Fields labeled as **N** acquire a null value by default if a value is not provided, all other fields apply a mandatory constraint.

reflected in the data sampled [3]. Water samples may also come from different sources including, wells, canals, and estuary. Even when two sampling locations are close to each other, it is of crucial to identify the water source as well.

The buoy data follows a spatio-temporal format, both location of the buoy and time of data collection must be taken into consideration when integrating the data with the information gathered from external sources. Another major challenge in data integration is the different frequencies in which the data is collected. The buoy sensor data is also a continuous time-series, meanwhile most of the relevant public data available follow a more discrete format with varying sampling frequencies. The most commonly used approach, re-samples both the continuous and discrete sets to compatible sampling frequencies; however, this can result in the loss of some valuable information [15]. Spatial interpolation and temporal interpolation methods are accessible by the domain expert who will decide the best approach considering the study's objectives and the data available [16].

Our proposed platform automates the retrieval of relevant water quality data from well-known open sources of hydrologic information—including South Florida Water Management District's (SFWMD) DBHYDRO, U.S. Geological Survey's (USGS) National Water Information System (NWIS), and Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI)'s HydroShare. The platform is built with the flexibility to easily extract the relevant data from communication endpoints developed to connect to the previously introduced open data sources. Such endpoints include dbhydroR [17], an R interface to DBHYDRO, a database that stores hydrologic, water quality, and hydro-geologic data collected in South Florida, and contains additional information about the location and context of where and the methods applied for data sampling. Using the location of the buoy, the nearest designated testing sites sampling the same water source are selected from DBHYDRO. Similarly, USGS allows the automated retrievals of their water data through a web service [18] and HydroShare provides a public RESTful web-service API [19] allowing the platform to use and integrate their data discovery functionalities.

D. Monitoring

As described in section IV-B, not all anomalies in the data result from false readings. Further studies are required to discover the reason behind the unusual values from valid yet anomalous behavior. In most cases, the user might decide to perform other sampling methods at the buoy's sampling site, such as the traditional and manual grab sample approach, and compare the measurements for further validation that the readings are correct. The user can also conduct studies about possible changes that might have taken place at the sampling site. The platform assists the study process by automating the creation of some essential analytical reports, including daily/monthly trends, yearly average, and changes in measurement through time.

V. Applications

Time-based information communication, whether in the form of hand sketches or volumetric isosurfaces, has always been an essential aspect of sciences and scientific information dissemination [20]. Visualization methods and platforms have evolved to not only allow the scientists to record their findings more accurately and efficiently but also

Figure 4. Mobile-based Augmented Reality application to monitor the water buoys.

to convert the same data into different types, ranging in two and three-dimensions, to better communicate with their peers and the public. Thus, information visualization has been an important component, especially in environmental sciences, for communicating findings with the stakeholders and policymakers to mobilize environmental policy change. In this context, we have developed two engaging near-real-time visualization methods to reach a diverse range of audiences and achieve maximum impact.

A. A Web-based Visualization Application

The first method is a web-based visualization built on HydroSphere's cloud-based capabilities discussed in section IV and accessible from the CREST's public website through a page dedicated to the Buoys' research. Here, the visitor finds detailed information about the research goals, sensors, as well as the specific location of each buoy with a link to lunch the visualization page. The visitor can visualize the data in chart view or table view. Each visualization mode is capable of filtering data based on the duration of acquisition and the parameters. Although both methods are using the same data, the chart view allows for an interactive histogram approach to plotting parameters to highlight trends and co-relations. The table view, on the other hand, allows the visitor to focus more on the numerical data. These web-based visualizations are both popular with the CREST scientists and the extended community.

B. Augmented Reality (AR) Mobile Application

An AR experience integrates the digital and physical worlds by seamlessly overlaying digital content on the surrounding environment. This technology renders AR experiences immersive, interactable, entertaining, and location-specific [21]. Building on these attributes, we have created a novel mobile AR experience that aims to not only display the near-real-time data from the buoys but also educate the public about the sites of data collection as well as the buoy technology for water quality monitoring. For that, the AR platform is built with the Unity game engine using the AR Foundation package to allow standalone application deployment for both Android and iOS platforms. The AR experience begins with the user launching the mobile application and pointing the device's external camera toward a planar surface, which triggers the appearance of a highresolution three-dimensional topographic model of the City of Miami and its peripheries. By incorporating strategic User Interface (UI) design, the UI buttons on the screen, enable the user to display the location of the buoys as three-dimensional icons and highlight the placement of the buoys to the varied conditions of the city, as shown in Figure 4. Furthermore, the user can click the buoy icon and zoom into its location, as well as pan and orbit around the icon to understand the nuances of its immediate context. Another button allows the user to display near-real-time data above the buoy icons, and using a slider go through the archived data for each specific buoy. Complimentary to this data visualization feature is an interactive histogram chart that allows the user to select different parameters and plot an interactive graph for analyzing data trends. In order to inform the users about the buoy technology, a button displays a three-dimensional model of the buoy where the user can see and interact with different components of the mechanism and get a sense of its multiple sensor types and their relevance to water quality monitoring. This AR application has been play-tested during two recent major CREST events, which allowed us to gauge its engagement and educational significance. Future work on this application will entail the following: (1) Incorporating a systematic method for user evaluation to maximize its impact on environmental literacy, (2) In addition to reporting data, suggest solutions to users for bettering their local water quality. We believe that this trajectory will assist us in impacting local policies.

VI. Future Directions

Our platform will be expanded to include more advanced techniques that serve to better manage and validate the data collected by the buoys.

A. Training a Smart Outlier Detection

The data processing and outlier detection techniques introduced in sections IV-B and IV-C can be further extended by training models to make smarter decisions about the data. When a domain expert distinguishes the noisy measurements from the pre-labeled anomalous data points, a supervised model can be trained to automatically learn the differences from the decisions made by the expert and make more accurate predictions.

B. Handling Missing Values

When noisy data points are identified and removed, the challenge now is how to compensate for the missing values. In addition, there are also other causes of missing values specific to the buoy's sensed data unrelated to the processing step. In some occasions, the buoy might need to reduce the frequency of data transmissions due to power and connection issues. Buoys may also be temporarily pulled from the sampling site to protect them from forecasted events, such as Hurricane Dorian in September 2019 as demonstrated in Figure 1. Filling the missing gaps in the data guarantees obtaining a continuous time-series. There is no general solution to handle missing values [22], and further studies must be conducted to identify the best approach.

VII. Conclusion

Wireless sensor networks have become indispensable in the field of water quality monitoring to tackle the complex issues of environmental contamination. This paper introduces a first-of-its-kind customized system that integrates different sensors into a relatively-compact and portable buoy designed for water quality monitoring. A data management platform is then proposed to support the research conducted for the near-real-time buoy sensor data. To the best of our knowledge, most previously proposed data management platforms developed for water quality monitoring have mainly focused on facilitating data storage, retrieval, and offering some straightforward visualizations of water quality data. The proposed platform aims to address several the challenges concerning data management, processing, integration, and monitoring. More specifically, the platform provides, (1) a persistent storage to safely backup the buoy raw data before transformations and processing techniques are applied, (2) a web service to allow for easy on-demand retrieval of the processed data, (3) data integration methods to reconcile the near-real-time data from the buoys with data from several public sources, (4) monitoring the trends and possible peaks in the data that might need further analysis. We further demonstrate the applications, that have resulted from such a platform, developed to reach a wide range of audiences. And finally, we highlighted the next steps in this study.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. HRD-1547798. This NSF Grant was awarded to Florida International University as part of the Centers for Research Excellence in Science and Technology (CREST) Program. This is contribution number 973 from the Southeast Environmental Research Center in the Institute of Environment at Florida International University.

References

[1] C. G. Panayiotou, D. Fatta, and M. P. Michaelides, "Environmental monitoring using wireless sensor networks," *Technical Report*, 2005.

[2] Environmental Protection Agency (EPA), "Nutrient Pollution: The Issue," Feb 2019. [Online]. Available: https://www.epa.gov/nutrientpollution/issue

[3] S. Mitra, P. Patnaik, and B. B. Kebbekus, *Environmental chemical analysis*. CRC Press, 2018.

[4] K. S. Adu-Manu, C. Tapparello, W. Heinzelman, F. A. Katsriku, and J.-D. Abdulai, “Water quality monitoring using wireless sensor networks: Current trends and future research directions,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 1, pp. 1–41, 2017.

[5] E. Skarbøvik and R. Roseth, “Use of sensor data for turbidity, ph and conductivity as an alternative to conventional water quality monitoring in four norwegian case studies,” *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, vol. 65, no. 1, pp. 63–73, 2015.

[6] J. S. Horsburgh, J. Caraballo, M. Ramírez, A. K. Aufdenkampe, D. B. Arscott, and S. G. Damiano, “Low-cost, open-source, and low-power: But what to do with the data?” *Frontiers in Earth Science*, vol. 7, no. 67, p. 1, 2019.

[7] H. B. Glasgow, J. M. Burkholder, R. E. Reed, A. J. Lewitus, and J. E. Kleinman, “Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies,” *Journal of experimental marine biology and ecology*, vol. 300, no. 1-2, pp. 409–448, 2004.

[8] J. S. Horsburgh, D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine, and T. Whitenack, “An integrated system for publishing environmental observations data,” *Environmental Modelling & Software*, vol. 24, no. 8, pp. 879–888, 2009.

[9] A. Samourkasidis, E. Papoutsoglou, and I. N. Athanasiadis, “A template framework for environmental timeseries data acquisition,” *Environmental Modelling & Software*, vol. 117, pp. 237–249, 2019.

[10] R. Richards, *Representational State Transfer (REST)*. Berkeley, CA: Apress, 2006, pp. 633–672.

[11] C. Severance, “Discovering javascript object notation,” *Computer*, vol. 45, no. 4, pp. 6–8, 2012.

[12] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413–422.

[13] ——, “Isolation-based anomaly detection,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012.

[14] N. R. Council *et al.*, *Catalyzing inquiry at the interface of computing and biology*. National Academies Press, 2006.

[15] G. H. Orcutt, H. W. Watts, and J. B. Edwards, “Data aggregation and information loss,” *The American Economic Review*, vol. 58, no. 4, pp. 773–787, 1968.

[16] L. Li and P. Revesz, “Interpolation methods for spatio-temporal geographic data,” *Computers, Environment and Urban Systems*, vol. 28, no. 3, pp. 201–227, 2004.

[17] J. Stachelek, “dbhydroR: An R package to access the DBHYDRO environmental database,” 2016.

[18] United States Geological Survey (USGS), “USGS water data for the nation automated retrievals.” [Online]. Available: https://waterdata.usgs.gov/

[19] D. G. Tarboton, J. S. Horsburgh, D. R. Maidment, T. L. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine, and T. Whitenack, “Development of a community hydrologic information system,” in *18th World IMACS Congress and International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, MODSIM09*, 2009, pp. 988–994.

[20] M. C. Nisbet, L. Kahlor, and P. A. Stout, *Communicating Science: New Agendas in Communication*. Routledge, 2010.

[21] S. C.-Y. Yuen, G. Yaoyuneyong, and E. Johnson, “Augmented reality: An overview and five directions for AR in education,” *Journal of Educational Technology Development and Exchange (JETDE)*, vol. 4, no. 1, p. 11, 2011.

[22] F. M. Shrive, H. L. Stuart, H. Quan, and W. Ghali, “Dealing with missing data in a multi-question depression scale: a comparison of imputation methods,” *BMC Medical Research Methodology*, vol. 6, pp. 57 – 57, 2006.