Pacific Northwest NATIONAL LABORATORY

PNNL-32294

# Data Fusion to Enhance Quality Control and Analysis with Instruments at the Marine and Coastal Research Laboratory

October 2021

Jonathan Whiting
Emma Cotter
James McVey
Nolann Williams
Shon Zimmerman
John Vavrinec
Erin Walters

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
*operated by*
BATTELLE
*for the*
UNITED STATES DEPARTMENT OF ENERGY
*under Contract DE-AC05-76RL01830*

**Printed in the United States of America**

**Available to DOE and DOE contractors from the**
**Office of Scientific and Technical Information,**
**P.O. Box 62, Oak Ridge, TN 37831-0062;**
**ph: (865) 576-8401**
**fax: (865) 576-5728**
**email: reports@adonis.osti.gov**

**Available to the public from the National Technical Information Service**
**5301 Shawnee Rd., Alexandria, VA 22312**
**ph: (800) 553-NTIS (6847)**
**email: orders@ntis.gov <https://www.ntis.gov/about>**
**Online ordering: http://www.ntis.gov**

PNNL-32294

# Data Fusion to Enhance Quality Control and Analysis with Instruments at the Marine and Coastal Research Laboratory

October 2021

Jonathan Whiting
Emma Cotter
James McVey
Nolann Williams
Shon Zimmerman
John Vavrinec
Erin Walters

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

# Abstract

Deploying environmental monitoring instruments in the marine environment can be challenging, facing challenges around device survivability, biofouling and corrosion, and consistent data collection. This project explores the use of data fusion – the process of integrating multiple data sources to produce more consistent, accurate, and useful information – to build a consistent long-term monitoring system at the Marine and Coastal Research Laboratory (MCRL) in Sequim, Washington. Unused instruments that had been acquired from past projects were inventoried and deployments planned on the MCRL pier and floating dock. A total of 8 instruments were deployed including a tide gauge, hydrophone, acoustic Doppler current profiler (ADCP), photosynthetically active radiation (PAR) sensors, meteorological station, and three water quality sensors. Deployments were planned to be well-protected around the pier structure and a maintenance schedule was created for cleaning and recalibration. An automated data pipeline was created to aggregate data on edge computers that push data to Amazon Web Services (AWS) cloud storage every 15 minutes, performing automated quality control and data transformations using the Time Series Data Analytical Toolkit (TSDAT). Continued efforts are underway to maintain this system into the future, take a data-driven approach to maintenance scheduling, improve the reliability of the system, and share the data with a variety of end-users.

# Acronyms and Abbreviations

| | |
|---|---|
| ADCP | acoustic Doppler current profiler |
| API | application programming interface |
| ARM | Atmospheric Radiation Measurement |
| AWS | Amazon Web Services |
| C | celcius |
| CDOM | colored dissolved organic matter |
| CTD | conductivity, temperature, and depth |
| dB | decibel |
| DOE | Department of Energy |
| FTP | file transfer protocol |
| g | gram |
| GB | gigabyte |
| GFCI | ground-fault circuit interrupter |
| HTTPS | hypertext transfer protocol secure |
| Hz | hertz |
| IoT | internet of things |
| L | liter |
| m | meter |
| MCRL | Marine and Coastal Research Laboratory |
| NAVD88 | North American Vertical Datum of 1988 |
| NetCDF | Network Common Data Form |
| PA | pascal |
| PAR | photosynthetically active radiation |
| PCA | principal component analysis |
| pH | potential of hydrogen |
| PNNL | Pacific Northwest National Laboratory |
| ppb | parts per billion |
| RPi | Raspberry Pi |
| s | second |
| S | siemens |
| SQL | Structured Query Language |
| TSDAT | Time Series Data Analytical Toolkit |
| UPS | uninterruptible power supply |
| UV | ultraviolet |

# Contents



# Figures

# Tables

# 1.0 Introduction

The Pacific Northwest National Laboratory's (PNNL's) Marine and Coastal Research Laboratory (MCRL), formerly known as the Marine Science Laboratory, was purchased by Battelle in 1966 and has performed decades of environmental research in areas of chemistry, wetland and coastal ecology, marine biology, and remote sensing. The lab is located next to the water on the entrance to Sequim Bay, Washington and regularly pumps seawater directly into laboratory facilities. In recent years, MCRL has been recognized as a test bed for marine instrumentation, particularly around marine energy applications. Yet despite all the advanced research, the laboratory has never established a long-term consistent monitoring network of Sequim Bay.

Instrumenting the marine environment presents several unique challenges. Debris can easily damage or destroy instruments, particularly during storms, so instruments must be protected against such hazards. Biofouling is the accumulation of plants, algae, and animals that can degrade measurements, requiring regular cleanings to physically remove the buildup and reapply substances like zinc oxide to slow future growth. Growth rates differ depending on the time of year, so cleaning intervals are not regular and depend on light, temperature, and nutrient availability at different times of the season. Cleaning intervals can be further refined with a data-driven approach.

This project explores applying data fusion to the challenges of the marine environment, to enable more effective long-term monitoring. Data fusion is a concept that has been around for decades (Chair et al. 1986, Hall and Llinas 1997) and has been researched at PNNL. The multi-lab Atmospheric Radiation Measurement (ARM) user facility explored the concept for atmospheric measurements and PNNL's SMART initiative is just beginning to explore the topic for subsurface measurements. Outside of PNNL, data fusion has been applied to ocean research mostly around integrated sensors in a single platform. This project integrates data from traditionally separate sensors and develop techniques that can be transferred to other marine applications. Building from other PNNL programs, the application of data fusion to the marine environment will be unique for the Department of Energy (DOE) system.

Two distinct goals around data fusion were (1) application to quality control and (2) to address a complex analysis as proof of concept. Quality controls include comparing data that should be physically correlated, or two instruments that monitor the same parameter, and tracking when they deviate from expectations. For example, a storm may cause an anomaly that exceeds expected limits, but if the anomaly is detected on multiple sensors, then it implies an extreme event rather than sensor degradation. Comparing related sensors may also better track sensor creep and can inform cleanings and recalibrations. The proposed proof-of-concept analysis was to determine whether winter storms or vessel traffic contribute more to channel erosion to Sequim Bay, making use of 5 different sensors. The meteorological station identifies storms, tide gauge informs flood and ebb cycles, hydrophone identifies vessel passage and estimates channel erosion with sediment noise, acoustic Doppler current profiler (ADCP) determines current speeds, photosynthetically active radiation (PAR) sensors estimate channel erosion as turbidity.

# 2.0 Methods – Instrumentation

An inventory was taken of instruments previously purchased at MCRL that were not designated for use. Working with experienced researchers who are domain experts with specific sensors, the list was narrowed to key sensors that collect the most essential information for addressing the research questions and providing baseline data for the Sequim Bay channel. These same researchers advised on the positioning and orientation of the instruments, sampling frequency, calibration procedures, maintenance schedules and informed the maintenance schedules. The following provides details on the various deployed sensors:

## 2.1 Tide Gauge – Nile YSI Micropilot FMR51

**Installation Date:** 4 March 2021

**Location and Orientation:** The tide gauge attached to the northeast side of the pier near the ramp to the floating dock, pointing down at the surface of the water (Figure 1).


Figure 1. Location and picture of the deployed tide gauge.

**Collected Parameters:**

- Elevation referenced to NAVD88 (m)

**Frequency of Collection:** 1 reading every 5 minutes

**Calibration Procedures:** Rough calibration of the sensor can be performed in a quick test by using a measuring device such as a laser range finder held next to the sensor and comparing the return distance to the datastream from the sensor.

When there is need for higher resolution calibration, the sensor should be removed from the mounting bracket and fixed to a mount that can be aimed at a fixed, flat target. Multiple range tests are recommended for a variety of ranges within they typical operating range of the sensor, i.e. 0.5 m, 1.0 m, 1.5 m, 2.0 m. A reference sensor can be used to determine the calibration distance or precise measurements can be made by hand instead. For a more detailed procedure see the section on range calibration in Heitsenrether and Davis (2011).

**Maintenance Required:** Maintenance for this device is not regular. If anomalous readings are received, check the sensor head for spiderwebs or other debris that may be blocking the signals. For more advanced troubleshooting please refer to the manufacturer's instructions.

## 2.2 Hydrophone – Ocean Sonics icListen SC2-ETH-X2

**Installation Date:** 10 September 2021

**Location and Orientation:** Northeast pier piling, 2 ft above the seafloor, sensor directed “up” (Figure 2).

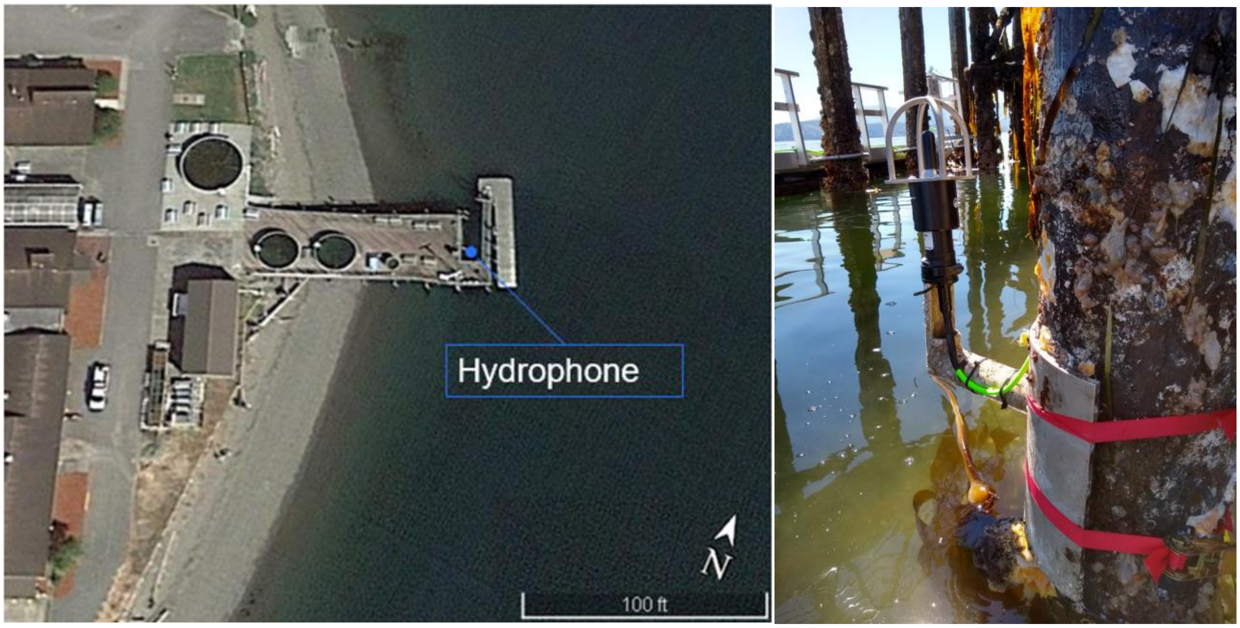


Figure 2. Location and picture of the deployed hydrophone. This photo was taken at an extreme low tide; the hydrophone is typically subsurface.

**Collected Parameters:**

- Averaged frequency spectra between 0 and 204 kHz (dB Ref re 1 V is 120, dB ref re 1 uPa is 87)
- Conditions (humidity, temperature)

**Frequency of Collection:** One 5-minute averaged reading every 15 minutes.

**Calibration Procedures:** Approximately every 2 years, the hydrophone needs to be calibrated for both high and low frequencies. Low frequency calibration can be done by Ocean Networks Canada, while high frequency can be calibrated at PNNL Richland. A pistonphone calibration tool can be purchased to confirm proper calibrations, providing additional confidence in the measurements.

**Maintenance Required:**

1. Clear biofouling and reapply zinc oxide paste to transducers every 1-3 months
2. During cleaning or maintenance, use caution to not damage the transducer element as it is relatively sensitive and easy to damage
3. A protective shield (shown in Figure 2) is used to protect the transducer and can be removed for cleaning but should always be replaced prior to redeployment.

## 2.3 ADCP – 400 kHz Nortek 2D Horizontal Profiler

**Installation Date:** 8 September 2021

**Location and Orientation:** Southeast pier piling, 5 ft above the seabed, facing true East (actual: 82 degrees CW from true N) (Figure 3).

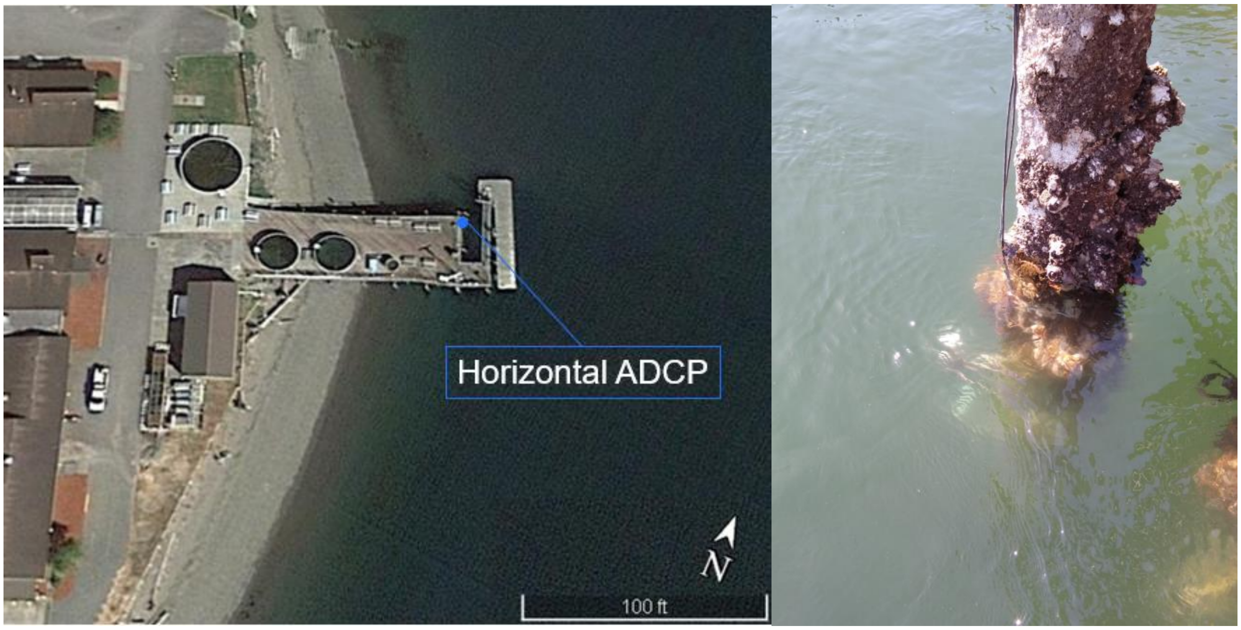


Figure 3. Location and picture of the deployed ADCP (submerged).

**Collected Parameters:**

- Horizontal velocity flow direction in each range bin, 6-98 m range bin sizes (degrees)
- Horizontal velocity magnitude in each range bin, 6-98 m range bin sizes (m/s)
- Positioning (pitch, roll, heading, amplitude)
- Environmental covariates (pressure, water temperature, depth)

**Frequency of Collection:** One 5-minute average profile every 15 minutes

**Calibration Procedures:** On a yearly basis, repeat the following steps:

1. In Nortek AWAC deployment software - stop recording
2. Remove instrument from water
3. Check connections are water tight, apply new grease to barrel connector pins
4. Set Pressure Offset: While instrument is in air, set the offset so that the pressure sensor reads "0.0" (dbar)
5. Compass Calibration: Mount the ADCP as it will sit in the sleeve to the aluminum bracket, set it on a cart and move it out to the middle of the pier, away from electronics/metal. Set the ADCP so that it sits flat (need a chock block or the like) and follow the steps in the compass calibration wizard in the software.
6. Set the instrument clock to the PC clock.

**Maintenance Required:**

1. Clear biofouling and reapply zinc oxide paste to transducers every 1-3 months
2. Replace zincs on aluminum bracket as necessary

## 2.4 PAR Sensors – Li-COR Li-193 with dedicated LI-1400 datalogger

**Installation Date:** 7 March 2021 (note: could not automate data collection due to age of logger, so data are offloaded manually)

**Location and Orientation:** Southeastern corner of the pier attached to a pile at elevations of 1.149 m and -2.149 m relative to NAVD88; sensors directed "up" (Figure 4).

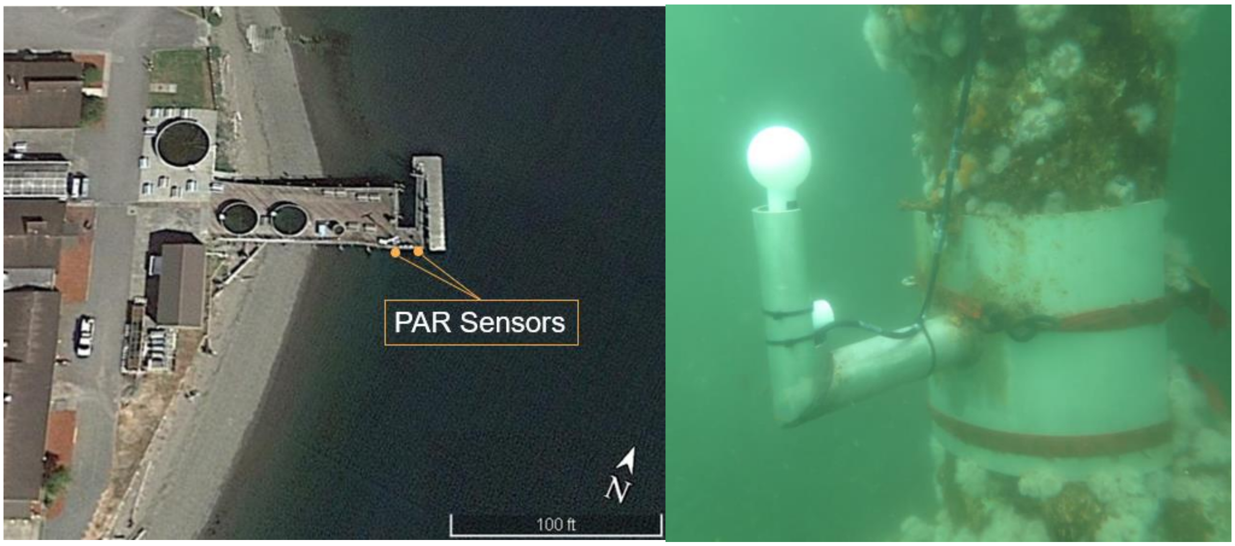


Figure 4. Location and picture of the deployed PAR sensors.

**Collected Parameters:**

- Photosynthetically active radiation flux density (µmol of photons m-2 s-1)

**Frequency of Collection:** Average photosynthetic photon flux density every 15 minutes.

**Calibration Procedures:** Return to Li-COR for manufacturer recalibration every 2 years unless noticeable scratches or damage appear in the bulb surface, in which case the bulb must be replaced (LI-COR 2006).

**Maintenance Required:** The PAR sensors are very sensitive to algal buildup that results in a sharp decrease in the recorded PAR flux density. They should be frequently cleaned with a microfiber cloth. Cleaning can occur during maintenance dives for the facility if time permits. Otherwise, additional dives will need to be scheduled monthly during the winter and twice monthly during the summer with the potential for additional cleaning as required. For more significant biofouling, devices can be removed and cleaned with a mild detergent or vinegar. DO NOT use alcohol, organic solvents, abrasives, or strong detergents.

## 2.5 CTD sensor – Sea-Bird SBE 37-SMP-ODO MicroCAT

**Installation Date:** 5 August 2021

**Location and Orientation:** The CTD sensor is attached to a protective metal cage suspended from the floating dock, attached to a pulley for ease of maintenance. Orientation does not matter as this is a point sensor (Figure 5).


Figure 5. Location and picture of the deployed water quality sensors.

**Collected Parameters:**

- Conductivity (mS/m)
- Water Temperature (°C)
- Dissolved Oxygen (mg/L)

**Frequency of Colletion:** 1 reading every 5 minutes

**Calibration Procedures:** Manufacturer recommends factory calibration when needed.

Conductivity Sensor: Recommend recalibration before and after deployment, but particularly when the cell has been exposed to contamination by oil or biological material.

Temperature Sensor: Should only drift a few thousandths of a degree in the first year and less after. Only recalibrate if there are indications that temperature is faulty.

Dissolved Oxygen: Primary issue with optical sensors is fouling of the optical window. Recommend recalibration before and after deployment, but particularly when the cell has been exposed to contamination by oil or biological material. The sensor film should be recalibrated when enough samples are taken to fill the memory of the device (~300,000 – 500,000 samples).

**Maintenance Required:** Every 1-3 months, clear biofouling from this instrument and clean the connectors, cable, and dummy plug per manufacturer instructions (Sea-Bird 2021). Replace parts as necessary. This sensor can be cleaned without dive team support because the sensor is suspended from the floating dock.

## 2.6 CDOM sensor – ECO Triplet-w

**Installation Date:** 5 August 2021

**Location and Orientation:** The CDOM sensor is attached to a protective metal cage suspended from the floating dock, attached to a pulley for ease of maintenance. Orientation does not matter as this is a point sensor (see previous Figure 5).

**Collected Parameters:**
- Chromophoric Dissolved Organic Matter (CDOM) (ppb)
- Chlorophyll (µg/L)
- Phycoerythrin (ppb)

**Frequency of Colletion:** 1 reading every 5 minutes

**Calibration Procedures:** Calibration should be performed by the manufacturer, but it is recommended that field characterization on fluorometers be performed occasionally as a field characterization.
1. Get a solution of known concentration.
2. Measure and record the solution output using the sensor.
3. Measure and record sensor dark counts.
4. Determine sensor scale factor as: Scale factor = x / (output – dark).
5. Determine solution concentration as: Concentration = (output counts – dark counts) * Scale factor.
6. Store scale factor and dark counts in the device file for the sensor, the internal memory of the sensor or both.

**Maintenance Required:** Every 1-3 months, clear biofouling from this instrument and use soapy water to clean oils from the optical face of the sensor. Replace the wiper or wiper motor as necessary (Sea-Bird 2017). This sensor can be cleaned without dive team support because the sensor is suspended from the floating dock.

## 2.7 pH sensor – Onset mx2501

**Installation Date:** 13 August 2021 (note: could not automate data collection due to device configuration, so data are offloaded manually)

**Location and Orientation:** The pH sensor is attached to a protective metal cage suspended from the floating dock, attached to a pulley for ease of maintenance. Orientation does not matter as this is a point sensor (see previous Figure 5).

**Collected Parameters:**
- Potential of hydrogen (pH), unitless scale of 0 to 14

**Frequency of Colletion:** 1 reading every 5 minutes

**Calibration Procedures:** The sensors should be calibrated with a 4,7,10 pH solution. The app used to download the data will provide a notification when calibration is necessary (Onset 2021).

**Maintenance Required:** Every 1-3 months this instrument should be cleared of biofouling, using mild dishwashing soap on the housing. The pH electrode should be inspected for deposits – scale and salt can be removed by soaking in 5% hydrochloric acid, oil can be removed by soaking in mid dishwashing soap. Add silicone-based grease to the outer O-rings. Replace parts from manufacturer as necessary. This sensor can be cleaned without dive team support because the sensor is suspended from the floating dock.

## 2.8 Meteorological Station – Assorted instruments

This station includes a Young Wind Monitor (anemometer), BaroVUE 10 Barometer, Hygrovue10 Temperature & RH, LI190SB Li-cor PAR Sensor, TE525 Rain Gauge Tipping Bucket, and a Campbell Scientific Datalogger CR1000X (Campbell Sci 2021a).

**Installation Date:** 27 September 2021

**Location and Orientation:** The meteorological station is attached to the southeast pier railing (Figure 6).

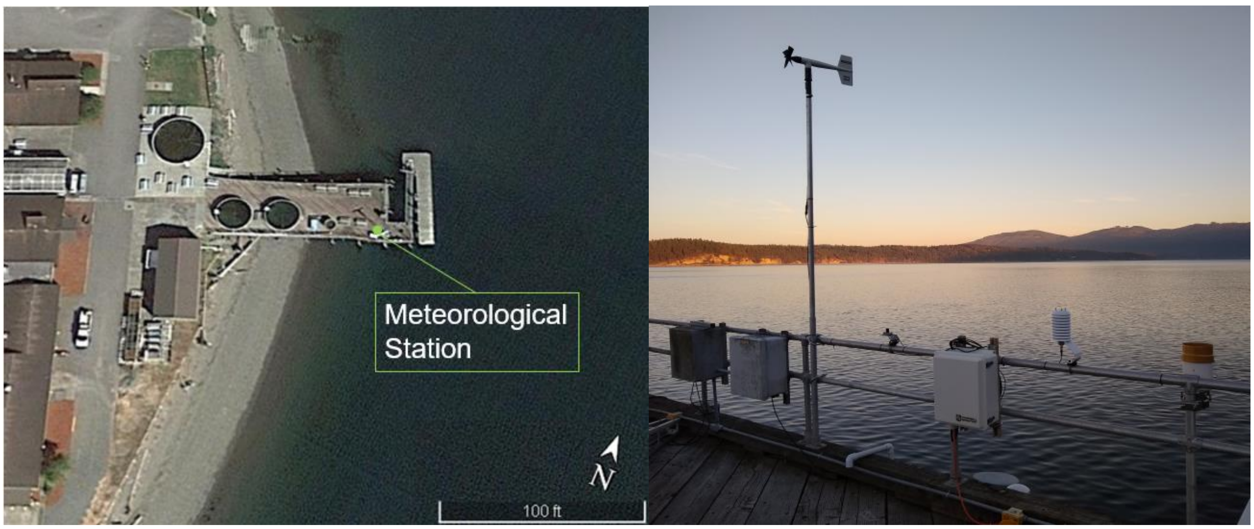


Figure 6. Location and picture of the deployed meteorological station.

**Collected Parameters:**

- Wind Direction (degrees)
- Average Wind Speed (m/s)
- Max Wind Speed (m/s)
- Barometric Pressure (Pa)
- Air Temperature (°C)
- Relative Humidity (percent)
- Photosynthetically active radiation flux density (µmol of photons m-2 s-1)
- Photosynthetically active radiation total flux (µmol of photons m-2)
- Rainfall (mm)

**Frequency of Collection:** 1 reading every 5 minutes. Windspeed values and PAR flux density are either an average across the 5-minute collection time, while total PAR flux and rainfall are sumed across the 5-minute collection time.

**Calibration Procedures:** Calibration depends on the specific instrument, as follows:

Young Wind Monitor (Campbell Sci 2020)

- Maintain bearings and potentiometer, no calibration needed
- May return to Campbell Sci for calibration/repair if needed

BaroVUE 10 Barometer (Campbell Sci 2021b)

- Replace sensor card annually (old one may be sent to Campbell Sci for recalibration)

Hygrovue10 Temperature & RH (Campbell Sci 2021c)

- Replace sensor element every two years or earlier if data becomes noisy (discard old elements – they cannot be recalibrated)

LI190SB Li-cor PAR Sensor (Campbell Sci 2015)

- Return to Campbell Sci every two (2) years for recalibration

TE525 Rain Gauge Tipping Bucket (Campbell Sci 2018)

- A field calibration check is advised every 12 months.
- *Field Calibration Check:*
  1. Secure a can or bottle that will hold at least 16 oz of water.
  2. Punch a very small hole in the bottom of the can or bottle. If it takes less than 45 minutes for 16 oz of water to run out, the hole in the can is too large.
  3. Place the can in the top funnel of the rain gage and pour 16 fluid ounces of water into the can. (A 16 oz soft drink bottle filled to within 2.5 inches of the top may be used for a rough field calibration. An exact volume will allow for a more precise calibration.)
  4. The following number of tips should occur: TE525, TE525MM 100 ± 3 TE525WS 57 ± 2
  5. Adjusting screws are located on the bottom adjacent to the large center drain hole. Adjust both screws the same number of turns. Rotation clockwise increases the number of tips per 16 oz. of water; counterclockwise rotation decreases the number of tips per 16 oz. of water. One half turn of both screws causes a 2% to 3% change.
  6. Check and re-level the rain gage.
- *Factory Calibration:*
  If factory calibration is required, contact Campbell Scientific to obtain an RMA

**Maintenance Required:** Maintenance depends on the specific instrument, as follows:

Young Wind Monitor

- Monthly: do a visual/audio inspection of the anemometer at low wind speeds. Verify that the propeller and wind vane bearing rotate freely. Inspect the sensor for physical damage.
- Replace the anemometer bearings when they become noisy, or the wind speed threshold increases above an acceptable level. The condition of the bearings can be checked by using the Propeller Torque Disc as described in the R.M. Young manual (see www.youngusa.com/products/7/).
- The potentiometer has a life expectancy of fifty million revolutions. As it becomes worn, the element can produce noisy signals or become non-linear. Replace the potentiometer when the noise or non-linearity becomes unacceptable. The condition of the vertical shaft (vane) bearings can be checked by using R.M. Young Vane Torque Gauge.
- NOTE: Campbell Scientific recommends factory replacement of the bearings and potentiometer. Refer to the Assistance page of this document for the procedure of acquiring a Returned Materials Authorization (RMA). Mechanically-adept users may choose to replace the bearings or potentiometer themselves. Instructions for replacing the bearings and potentiometer are given in R.M. Young manuals (www.youngusa.com/products/7/). A video that describes changing the bearings is available at: www.campbellsci.com/videos/wind-monitor-bearingreplacement

BaroVUE 10 Barometer

- Monthly: check humidity indicator inside Campbell Sci enclosure. Replace desiccants if needed
- Monitor quality metric. If < 4, replace BaroVUE 10 card.
- Annually: recommended to send sensor card to Campbell Scientific for calibration
- NOTE: sensor card is ESD sensitive

Hygrovue10 Temperature & RH

- Monthly: Check radiation shield to ensure it's free of dirt and debris. Clean with soap and water if needed, allow to dry before remounting. Check white filter on end of sensor for debris. Clean with distilled water or replace filter if needed.
- Annually: replace sensor element

LI190SB Li-cor PAR Sensor

- Monthly: Check leveling fixture is still level
- Monthly (or as needed): clean sensor with water and/or mild dishwashing soap. Vinegar may be used to remove hard water deposits. DO NOT use alcohol, organic solvents, abrasives, or strong detergents.

TE525 Rain Gauge Tipping Bucket

- Monthly: Check for and remove debris
- Annually: Field calibration check

# 3.0 Methods – Data Pipeline

One of the goals for this project was to create smarter quality controls, which is best done in real time. Therefore, the data collection was designed to occur near-real-time with goals to make it easy to maintain, scalable, and cheap. Design included instrument controls to manage collection and pushes to Amazon Web Services (AWS), conducting quality controls and data transformations before storage in AWS, and options for data accessibility. The overall architecture is summarized in Figure 7.

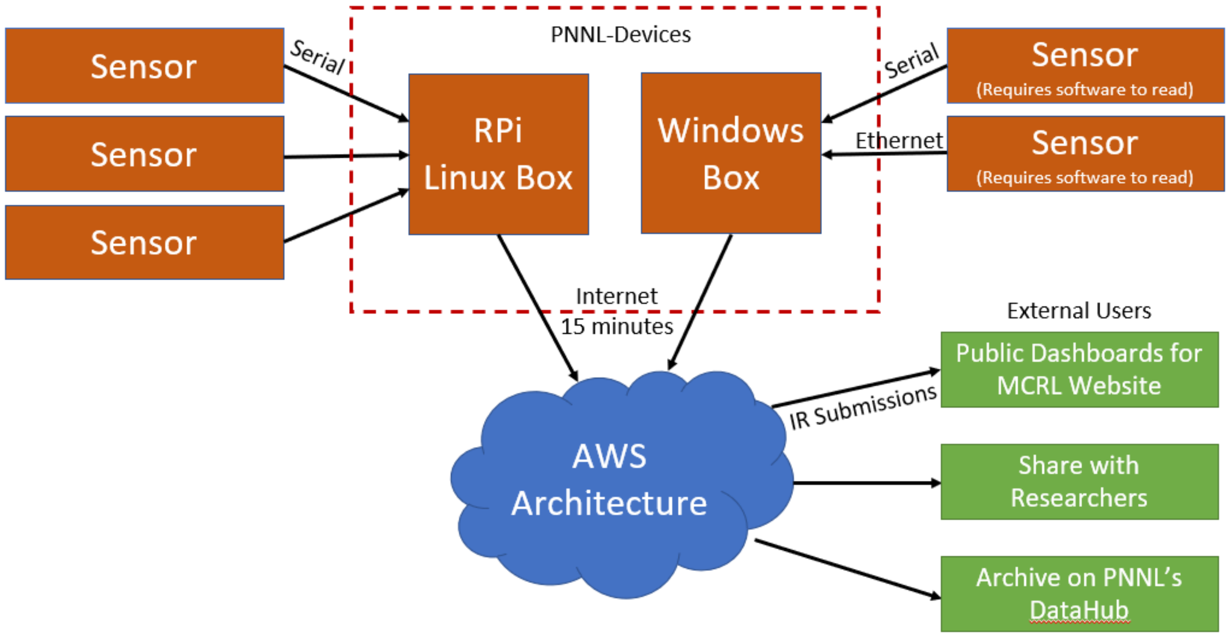


Figure 7. Data pipeline architecture including collection by sensors, controls with edge computers, pushes to AWS storage, and options for sharing data.

## 3.1 Instrument Controls

As instruments were deployed, a wired serial connection was connected to edge computers that control the sampling frequencies and the timing of data pushes to AWS. A Raspberry Pi (RPi) running Linux was used to control the tide gauge and various water quality instruments. Some instruments required specialized software to decipher output data, and this software often came native to a Windows operating system. While a virtual environment could have been stood up on the RPI, it was simpler to purchase a separate Windows box. A rugged fanless industrial mini PC running Windows was used to control the hydrophone, ADCP, and meteorological station. It should noted that the PAR sensors and pH sensor did not have options to connect directly with the edge computers and data offload needs to be done manually.

Both edge computers were configured to run using the cron command-line utility ("cron jobs") to manage sampling frequencies as specified above. Data readings are collated into a 15-minute file chunks per instrument and sent to AWS across Wi-Fi on the PNNL-Devices network. In case of lapses in Wi-Fi signal, files are staged in a folder until the upload to AWS is successful. At that point, the data file is also copied into a local backup folder where it is retained for 1 month for redundancy.

## 3.2 AWS Collection and Storage

The initial design decision was whether to use traditional physical storage or cloud storage through AWS. The following table was created to outline the pros and cons for each option (Table 1).

Table I. Pros and Cons of Cloud Storage versus Physical Storage.

| Cloud Storage (AWS) | Physical Storage |
|---|---|
| **Costs:** Pay for exact storage used, no up-front costs | **Costs:** High capital costs, eventually need replacement |
| **Backups:** Assured by AWS, no effort | **Backups:** Double the storage, requires management |
| **Upkeep:** Serverless, health monitoring | **Upkeep:** Hardware and software upgrades |
| **Upload:** Requires an internet connection, enables remote field deployments | **Upload:** Requires physical collection and upload to storage |
| **Size:** Designed for millions of devices sampling frequently, supports any storage amount | **Size:** Must predict future storage needs |
| **External Access:** No PNNL network restrictions, simple SQL query to share data (download, API), potential analysis within cloud | **External Access:** Some restrictions by PNNL network, some transfer options (e.g., FTP) |
| **Sensitive Data:** Not configured to support, but options available | **Sensitive Data:** Data can be protected |

The decision was made to go with cloud storage in an AWS architecture designed to be low cost and low maintenance. The AWS architecture was designed to emulate PNNL's Atmospheric Radiation Measurement (ARM) program, which has collected continuous atmospheric measurements from around the United States for nearly 30 years. The AWS architecture includes use of IoT (Internet of Things) Core, S3 storage, Lambda functions, and AWS Athena. This framework is serverless pay-per-use approach that does not include database maintenance and upkeep. While AWS is designed for billions of devices at second collection intervals, the cost for several instruments at 15-minute collection intervals is quite minimal, on the order of several hundred dollars per year.

IoT Core enables easy and secure connections between devices and the cloud. An HTTPS communication protocol was selected, using mutual authentication and end-to-end encryption to pass the data securely to AWS. This service currently offers a free tier of 2.25 million connection minutes and 500,000 messages per month. Current utilization is well within the free tier.

S3 storage is not a database but is essentially structured file storage. Several tiers are available, where costs become cheaper but queries take longer as you select more infrequent access tiers. The standard tier was selected for this pipeline because the quantity of data is relatively small and frequent queries are expected. Costs are currently at $0.023 per GB per month. Naming structure is important, as it partitions data for more effective querying after storage. Our pipeline uses the following structure after processing with Lambda functions: bucket/root/device/details-sampling-filetype/year/month/day (i.e., mcrl-data-processed/root/ctd/mcrl_pier-seabird_smp_1-5min.parquet/2021/09/20). The naming schema includes specific device details such as location and manufacturer, the sampling frequency, and date partitions below the file type, so we can query a specific day for a specific file type without scanning more data than necessary.

Lambda functions were used to manage quality control, data transformation, and storage. This approach is fast, serverless, scalable, and flexible, essentially using custom code to perform actions when triggered. The Time Series Data Analytical Toolkit (TSDAT) is an open-source Python framework created for processing data pipelines. TSDAT was created by researchers at PNNL for environmental monitoring applications using meteorological measurements, and has been adapted for use by MHKit, which is a toolbox of data pipeline tools for use in marine energy applications. The TSDAT template is designed with the option to test the pipeline locally before scaling up and deploying the pipeline code as a Lambda function on AWS. The pipeline code is triggered when the IoT Core places the files into an S3 bucket. The pipeline code performs several actions: (1) quality control flags are added for each detected parameter, for null values and exceedance of limits, at a basic level; (2) parameters are renamed or processed (e.g., convert units to standard values) as necessary; (3) relevant metadata is stored; and (4) reformatted data are stored in a new S3 bucket, while the original file is removed. Three output file types are stored: a duplicate of the raw files for backup, NetCDF files containing metadata information and multi-variate information for sensors like the ADCP and hydrophone, and parquet files store compressed single-dimension results that can be easily queried by AWS Athena. Lambda costs are $0.20 per 1M requests and $0.0166667 for every TB-second of processing.

AWS Athena is a simple query service using standard SQL. With datasets partitioned by date and stored in compressed formats, queries are very fast and cheap. Athena costs are $5.00 per TB of data scanned. Athena also has a connector to Tableau, which will be used to create live visualization dashboards for the most recent two-weeks of collected data, updated hourly. Results for longer datasets can be queried by Athena on request from other researchers or stakeholders. Finally, PNNL has a service called Data Hub, where we can archive old datasets to save on S3 costs and improve visibility, as Data Hub has a user-friendly interface for searching and generates a DOI and citation.

Public interactive visualization dashboards have been developed using Tableau. Following more thorough review by experts and potential instrument recalibrations with the manufacturers as necessary, these dashboards will be made available to the public. The visualizations were created from live Athena queries showing data for the last two weeks and be scheduled to update every hour. More frequent updates and longer datasets are possible, but Athena costs will be more significant.

# 4.0 Data Fusion and Analysis

This project was supported with laboratory directed research and development funds, which provides funding across a single fiscal year. Funds were received in November 2020 and the project ran through September 2021. Coronavirus challenges restricted travel, limited access to MCRL, and delayed purchase orders. These challenges in combination with limited staff availability led to some severe early delays in deploying devices. The tide gauge and PAR sensors were deployed in March 2021, but all other deployments were delayed until August and September 2021. There were limited data to analyze that missed the big winter storms important to the analysis, and there was little time to run advanced processing on the data. As such, this project was unsuccessful at addressing the proof-of-concept analysis around comparing channel erosion between winter storms and vessel traffic. Some quality controls were implemented with live data processing, but there was no time to fully implement cross-sensor calibrations as hoped. These research areas are still important and will continue to be pursued pending future opportunities.

Initial exploration of the data identified events that highlight the challenges of deployments in the marine environment. Figure 8 shows a two-week sample of dissolved oxygen data. There was a distinct moment where values stopped showing tidal signals and eventually flatlined at 0 mg/L, implying hypoxic conditions. However, this was likely caused by seaweed or some other obstruction producing invalid results, as the water quality instruments were cleaned on the evening of 15 September 2021 and values returned to normal. This event highlights the challenge of biofouling. This was a case of acute fouling, though data quality can slowly degrade over time as biofouling accumulates. Long-term analysis of the data can identify seasonal trends for biofouling and better inform maintenance schedules.

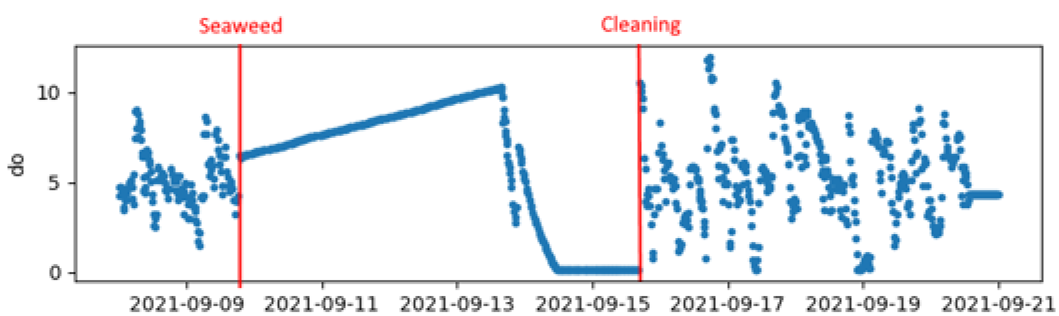


Figure 8. Dissolved oxygen data during 2 weeks in September when data failed until the water quality cage was cleaned.

To identify key trends between the data streams, a principal component analysis (PCA) was conducted. The first principal component accounted for 35% of the variance in the water quality data. This first principal component showed a strong correlation with water level data measured by the tide gauge (Figure 9), implying that water level is a key correlation in the water quality data. Water level identifies the tidal cycle, which flushes pollutants and circulates nutrients in the way, so it makes sense that water level would have such a strong correlation. This analysis demonstrates the type of analysis that can be done with co-located, co-temporal data and lays the framework for advanced quality control.

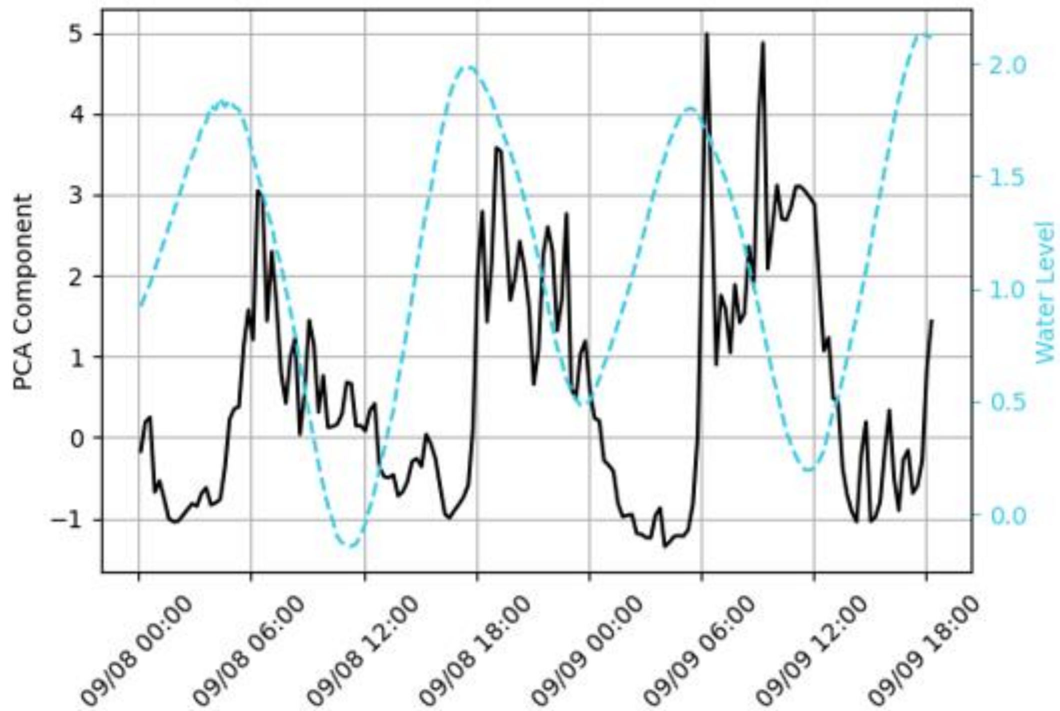


Figure 9. Plotting the first principal component of a PCA analysis (black) overlaid with water level (blue) to highlight the correlation.

# 5.0 Conclusions

This project built a robust baseline environmental monitoring program at the Marine and Coastal Research Laboratory. Eight environmental monitoring instruments were installed on the pier and floating dock near the entrance to Sequim Bay, Washington. Installations are intended for long-term deployment to characterize the baseline and seasonality of monitored parameters with the objective of better supporting various research efforts in Sequim Bay. Effort was made to create a robust system that is hardened against weather, reliable, and easy to maintain to reduce costs over the lifetime of the deployments. There were many lessons learned during the process.

The live data pipeline allows for the monitoring of incoming data. Occasionally, the data pipeline would start sending null values or stop sending data altogether. Troubleshooting identified common causes and some solutions were explored:

- The tide gauge would occasionally start throwing null values. Forcing a power cycle fixes the outputs, so a GFCI outlet that could be controlled by the RPi was installed and triggers a restart when the RPi detects five null values in a row.

- MCRL occasionally has brief power outages, and this would cause the edge computers to stop collecting data until manually restarted. An Uninterruptible Power Supply (UPS) was installed to maintain consistent power to the computers. The edge computers may temporarily lose their wifi connection and devices may temporarily lose power to take measurements, but the process continues as soon as power is restored.

- Wifi connection is occasionally lost, and collected data are not able to live push to AWS. This issue seems to be exaggerated by inclement weather. The edge computers have a staging folder that stores files locally until a connection is made and the files can be pushed to AWS. Wifi signal boosters are also being explored to make the pipeline more consistent. When the Wifi connection is lost, without physically checking the edge computers it is not possible to determine whether the sensor/computer is no longer recording data, or if data are simply being recorded locally.

Data analysis was limited because deployments were completed near the end of the fiscal year. However, the initial analysis identified strong correlations between water quality instruments and the tide gauge, laying the framework to further apply data fusion to quality controls in the future. These research efforts may continue in collaboration with TSDAT.

Funding has been secured from multiple projects and programs to maintain the deployments (Table 2). More sustainable funding options are being explored for long-term support of this system, such as infrastructure funds or additional laboratory investment for key deployments viewed as essential. Future efforts include ongoing cleaning and maintenance, regular scheduled recalibration, improvements to the framework to make the system more resilient and efficient, and new research partnerships. One planned improvement is to replace two of the problematic instruments. The PAR sensor logger does not support automation, so the logger will be replaced or bypassed with a direct connection to the RPi. The pH sensor only supports a Bluetooth connection for download that doesn’t support automation and requires manual recovery of the sensor, so there are plans to replace it with a Sea-Bird pH sensor that can integrate with the other Sea-Bird water quality instruments. There are also plans to explore the use of UV light to reduce biofouling, particularly on instruments where readings are sensitive to biofouling. Future efforts may also explore the side-by-side deployment of a cheap consumergrade instrument with an existing scientific-grade instrument, allowing cross-sensor calibrations while comparing the resiliency of cheaper sensors. Additional instruments being targeted for future deployments include measurements of CO2 in the water, depth profiles for water quality, and video to correlate events such as vessel passage. As additional instruments are suggested by partners, ongoing funding will be needed for continued maintenance of the deployment.

Table II. A list of projects planning to offer support for the continued maintenance of this system.

| Project | Interest |
|---|---|
| Ocean Observing Prize | The final phase of this national competition will involve testing prototypes at sea in Sequim Bay. Primary interest is in currents and CTD; minor interest in tide gauge, weather station, hydrophone |
| Testing Experience and Access to Marine Energy Research (TEAMER) | Offers testing capabilities at MCRL to marine energy technology developers. Interested in water PAR and temperature data, for upcoming laboratory projects drawing water from outside. |
| Triton | Researching and developing monitoring technologies and methods to understand potential environmental effects of marine energy devices. Interest in general water conditions for future work. |
| Strategic Environmental Research and Development (SERDP) project | Creating an underwater standardized test bed for assessing the performance of technologies that can detect, classify, and remediate unexploded ordinance. The array will be located elsewhere in Sequim Bay, so primary interest is in the meteorological station. |
| Cabled Research Array for the Blue Economy and Energy (CRABEE) | A test bed to support future deployments and demonstrations of community-scale tidal energy devices, blue technology, and oceanographic sensors and platforms. Primary interest is in the data pipeline. |
| WA State Demonstration Project | Demonstration of collecting renewable energy, powering a battery, and powering several end-uses. One end-use will be environmental monitoring, powering some of these devices. |
| Division Overhead | Recognizing the importance of this data collection effort to future research endeavors, support has been provided to maintain the system. |

Numerical modelers at PNNL have also built a model of Sequim Bay that can be calibrated and validated with measurements by this system. This creates opportunity to inform water conditions across the entire bay, while providing accuracy and ground-truthing for the model. Finally, the data is being made as open and available as possible, supporting future research endeavors. Long-term monitoring allowing the identification of long-term shifts in the baseline while providing expected conditions for researchers operating in Sequim Bay. Future additions to the instrumentation system should be targeted to specific research questions or else expand this baseline monitoring.

# 6.0 References


Campbell Sci. 2021a. *CR1000X Measurement and Control Datalogger.* Rev. 10/2021, Campbell Sci., Logan, UT, USA. https://s.campbellsci.com/documents/us/manuals/cr1000x-product-manual.pdf

Campbell Sci. 2021b. *BaroVUE10 Digital Barometer.* Rev. 8/2021, Campbell Sci., Logan, UT, USA. https://s.campbellsci.com/documents/us/manuals/barovue10.pdf

Campbell Sci. 2021c. *HygroVUE10 Temperature and Relative Humidity Sensor.* Rev. 7/2021, Campbell Sci., Logan, UT, USA. https://s.campbellsci.com/documents/us/manuals/hygrovue10.pdf

Campbell Sci. 2020. *Wind Monitor Series: 05103, 05103-45, 05106, 05108, 05108-45, and 05305*. Rev. 12/2020. Campbell Sci., Logan, UT, USA. https://s.campbellsci.com/documents/us/manuals/05103.pdf

Campbell Sci. 2018. *TE525-Series Tipping Bucket Rain Gages.* Rev. 7/2018, Campbell Sci., Logan, UT, USA. https://s.campbellsci.com/documents/us/manuals/te525.pdf

Campbell Sci. 2015. *LI190SB Quantum Sensor.* Rev. 2/2015, Campbell Sci., Logan, UT, USA. https://s.campbellsci.com/documents/us/manuals/li190sb.pdf

Chair, Z.; Varshney, P. 1986. Optimal Data Fusion in Multiple Sensor Detection Systems. IEEE Transactions on Aerospace and Electronic Systems, AES-22(1): 98-101. https://doi.org/10.1109/TAES.1986.310699

Hall DL, Llinas J. 1997. An introduction to multisensor data fusion. Proceedings of the IEEE, 85(1): 6-23. https://doi.org/10.1109/5.554205

LI-COR. 2006. LI-COR Underwater Radiation Sensors: Instruction Manual. 984-08307, LI-COR, Lincoln, NE, USA. https://www.licor.com/documents/285ymmolackqjidu9wao

Onset. 2021. *HOBO pH and Temperature Logger (MX2501) Manual*. 22511-P. Cape Cod, MA, USA. https://www.onsetcomp.com/files/manual_pdfs/22511-P%20MX2501%20Manual.pdf

Sea-Bird. 2021. *SBE 37-SMP-ODO MicroCAT C-T-ODO (P) Recorder*. Rev. 011, 9/2021, Sea-Bird, Bellevue, WA, USA. https://www.seabird.com/asset-get.download.jsa?id=54627862348

Sea-Bird. 2017. *User Manual: Eco Triplet-w*. tripetw170706, Rev. B, 7/2017, Sea-Bird, Bellevue, WA, USA. https://www.seabird.com/asset-get.download.jsa?id=54627862519

**Pacific Northwest National Laboratory**

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

***www.pnnl.gov***