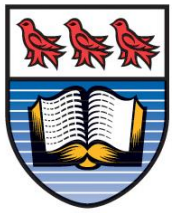

University of Victoria

Department of Physics and Astronomy

Interim Work Term Report

Spencer Plovie

spencerwplovie@uvic.ca

In partial fulfillment of the requirements of

the Physics and Astronomy Co-op Program

Spring 2023

Work Term #4 of 4

Junior Scientific Programmer

Confidentiality notice: This report is not confidential and may be shared with other co-op students.

This report will be handled by the UVic Co-op staff and will be read by one assigned report marker who may be a co-op staff member within the Physics and Astronomy Science Co-operative Educative Program, or a UVic faculty member or teaching assistant. The report will be retained and available to the student or, subject to the student's right to appeal a grade, held for one year after which it will be deleted.

I approve the release of this report to the University of Victoria for evaluation purposes only.

Supervisor's Signature: _____ Position: Senior Scientific Programmer Date: April 25, 2023

Supervisor's Name (print): Ben Biffard Email: bbiffard@oceannetworks.ca

Company: Ocean Networks Canada

Contents

I. Workplace Description

Ocean Networks Canada (ONC) is a data pipeline designed to deliver cabled, mobile, and community-based ocean data to the public. Established as a not-for-profit in 2007, the standardized data is freely available through ONC's Data Management and Archival System (DMAS), Oceans 3.0[1]. This archived and continuous data acquisition is available in a wide variety of formats.

Oceans 3.0 has tools to utilize and manage this dynamic database on both the developer and end-user sides and has several key aspects to its framework (Owens et al., 2022). There is an extensive and stable architecture composing the data acquisition and archival framework that allows for straightforward data versioning control, a detailed tracking system for all sensor metadata, quality assurance and quality control checks for all instruments complete with automated and manual testing procedures, data discovery and access tools including an Application Programming Interface (API), ability for interactive data handling including video annotations and more, persistent identifiers (PIDs) and citations to support the exchanging of data with external repositories, a system put in place to detect and respond to earthquakes, and of course a data product (DP) pipeline to provide data in numerous formats. The scientific programmer is primarily responsible for creating, maintaining, and improving these DPs working with the software development team.

Part of the work performed in this position was done using MATLAB, a programming language developed by MathWorks that is widely used by scientists and engineers for computation, visualization, and programming in data analysis and processing[2]. When a user goes to the Data Search page on the ONC website, they can select a device and choose the data product(s) they are interested in, then add this to their cart and 'checkout', where the raw data is externally retrieved and then processed through the DMAS framework and returned to the user as a compressed file, with a search ID associated with the parameters the user specified. Some of these searches rely on ONC's MATLAB-based framework to process the data product which, depending on the device chosen, can call a wide variety of functions designed to process data from different devices – fluorometers, seismometers, spectrometers, and much more.

In this report I will be giving a brief overview of the work completed during this term in the form of an interim report. This is not in the format of a final report, as the requirements have already been satisfied by two previous reports, confirmed by UVic co-op coordinator Jeremy Pearce.

II. Summary

Much of the work done during this term consisted of an assortment of feature developments, bugfixes, and regular maintenance and testing of ONC's data products. Some of the 'major' developments done during this term included the addition of citations[3] in the MATLAB DPs to store metadata information, creating several class-based unit tests for the testing and development of a function used to append PDF files together in the ONC DP repository, starting the process of creating a new AQUAscat AQUATEC[11] DP, and a continuation of last term's work in creating a soundscape metrics and hydrophone diagnostics DP in Python. The unit tests were created using a class-based framework[10] and were used to

comprehensively test the different aspects of the function's code, including types of user input, warnings and errors issued, and check that the output matches as expected. The AQUAscat device is used to observe profile of suspended sediment concentration and relies on backscattering/acoustic profiling methods for data acquisition. The work done for the soundscape metrics data product has been ongoing since August 2022, and in the past four months has consisted of refining the plots to be more consistent with ONCs standards (including adding more supplementary title information, reformatting into landscape with proper dimensions, adding gaps in the plot where data is missing, making the colour scheme colourblind-friendly, and more – see Figure 1 for the most current version of the project's progress), as well as working on the MATLAB script to run the Python file with the *pyrunfile* function, and then have this run as a weekly task through ONC's Task Management feature.

Some of the smaller features added include several new features for the netCDFs, adding a webservice login for specific on-demand near-live products, and adding more frequency band and orientation options for the seismometer channels[5]. The features added to the netCDFs included starting refactoring the MATLAB code to standardize all netCDF files with consistent and compliant attributes[6], as well as updating the netCDF attribution fields to account for the citation changes to the DPs that was previously mentioned.

Several miscellaneous bugfixes were also made to the MATLAB code. As a brief overview, some of these fixes included better handling of poorly formatted responses from the IRIS webservice[7] for seismometer DPs, fixing discrepancies in our code as we migrated from MATLAB version R2019a to R2022b, an attempted data rescue for an ASL echosounder[8] that was reporting the incorrect channel info by adding a default channel order override option for the device, improved plotting logic in MATLAB when data is alternating with NaN values (since line style plots will not show in this specific case, scatter plots should be used instead), and reducing the amount of unnecessary warnings in the code (especially if it slows down a search significantly). In addition, there was a case with the hydrophones where diverted data[9] was being returned as decimated data without the proper Low-Pass Filter (LPF) label on the files – these files were searched for and identified in ONC's data archive by assuming a maximum file size that downsampled audio data files would have and then transferred as a project to ONC's data stewardship team to properly label these diverted files. Another minor bugfix that was investigated was an apparent discrepancy between two file formats for an ASL echosounder device – stepping through the code and comparing corresponding searches for the two formats revealed that missing data is handled differently for .mat and ASL binary file types, and cannot produce the exact same result – when an invalid line of hex data is encountered, the code set up to create the .mat file product will fill any missing data with NaN values, while the binary file will simply skip over that timestamp.

There are basic tasks completed on a regular basis as part of the responsibilities of the scientific programmer. An important aspect of software development is regression testing. Regression testing is used to confirm that recent changes have not affected existing features in unexpected or adverse ways, and for ONC consists of a mixture of automated and manual tests. Apart from the regular monthly regression testing done for all features planning to be deployed in the next release, sometimes certain manual tests are performed to test a new option that had been added – in particular, during my work term there were changes made to the scalar data products, where min/max and min/max + average

processing options were added. Given that these new changes affect the QAQC flag[4] assignments in the product, the QAQC DP regression test procedure was used to ensure the behaviour of the new additions was as expected. In QA, a specific sensor with several hours of 'fake' data that contains a mixture of QAQC markers set up specifically for this sort of test case was used to create search IDs with various resampling types and periods. The products output from these searches were then compared against the expected flags to ensure the new code handles each case properly. On a similar vein, another important aspect to consider with new developments are test cases, which are test search IDs used to ensure that all DP features are functioning as we expect and are added to a list of searchIDs run every night in the QA environment, known as search automation. This can include what are known as lifetime searches, which are test cases that cover the entire data availability time range for a particular device. This is done as a way to 'stress test' the task machines that these searches run on and check their impact on the reliability of the system. Throughout any changes made to the data products, the corresponding documentation on the ONC external wikipedia[3] also must be kept up to date so users have an accurate understanding of the data they choose to download.

III. Discussion (and Acknowledgements)

Each work term completed at ONC has compounded my knowledge of software development and made me more confident in my coding abilities, as well as helped me become a more well-rounded individual. Having collaborated with a variety of teams has helped me learn how to convey technical information to others and, likewise, interpret information on subjects I am not well-versed in and pick out the relevant information to complete the task at hand. In addition, working on multiple projects at once has helped me with my time management and project management skills and effectively prioritize and switch between tasks. I have also had the opportunity to develop my problem-solving and critical thinking skills through the numerous technical issues summarized in the previous section. These skills have helped me identify the root cause of a problem and develop the appropriate solution at a much faster pace and helped me analyze and understand others' code more quickly. I have no doubt that these skills will be useful throughout my career.

I would like to take the time to thank my mentor, Dr. Ben Biffard, for his guidance while I worked here at ONC as a co-op student. I greatly appreciate his patience and understanding while I was learning new skills and for providing guidance when I made mistakes.

IV. References

D. Owens et al. *The Oceans 2.0/3.0 Data Management and Archival System*. Frontiers in Marine Science, Volume 9 (2022); https://www.frontiersin.org/articles/10.3389/fmars.2022.806452

Websites

[1] Ocean Networks Canada, *About ONC*, accessed 18 August 2022. Available at: https://www.oceannetworks.ca/about-onc/

[2] Wikipedia, *MATLAB*, accessed 18 August 2022. Available at: https://en.wikipedia.org/wiki/MATLAB

[3] Ocean Networks Canada, *Data Products Home*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/DP/Data+Products+Home

[4] Ocean Networks Canada, *Quality Assurance Quality Control*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/DP/Quality+Assurance+Quality+Control

[5] Ocean Networks Canada, *Seismometer Channel Option*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/DP/Seismometer+Channel+Option

[6] Ocean Networks Canada, *NetCDF*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/DP/NetCDF

[7] IRIS, *Data Services*, accessed April 24, 2023. Available at: https://ds.iris.edu/ds/

[8] Ocean Networks Canada, *ASL Acoustic Profiler Time Series*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/DP/24

[9] Ocean Networks Canada, *Hydrophone Data Acquisition and Diversion Mode*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/DP/Hydrophone+Data+Acquisition+and+Diversion+Mode

[10] MathWorks, *Ways to Write Unit Tests*, accessed April 24, 2023. Available at: https://www.mathworks.com/help/matlab/matlab_prog/ways-to-write-unit-tests.html

[11] AQUATEC, *AQUAscat 1000R*, accessed April 24, 2023. Available at: https://www.aquatecgroup.com/aquascat/aquascat-1000r

[12] Ocean Networks Canada, *Creating API Requests Using Data Search*, accessed April 24, 2023. Available at: https://wiki.oceannetworks.ca/display/O2A/Creating+API+requests+using+Data+Search

V. Appendix

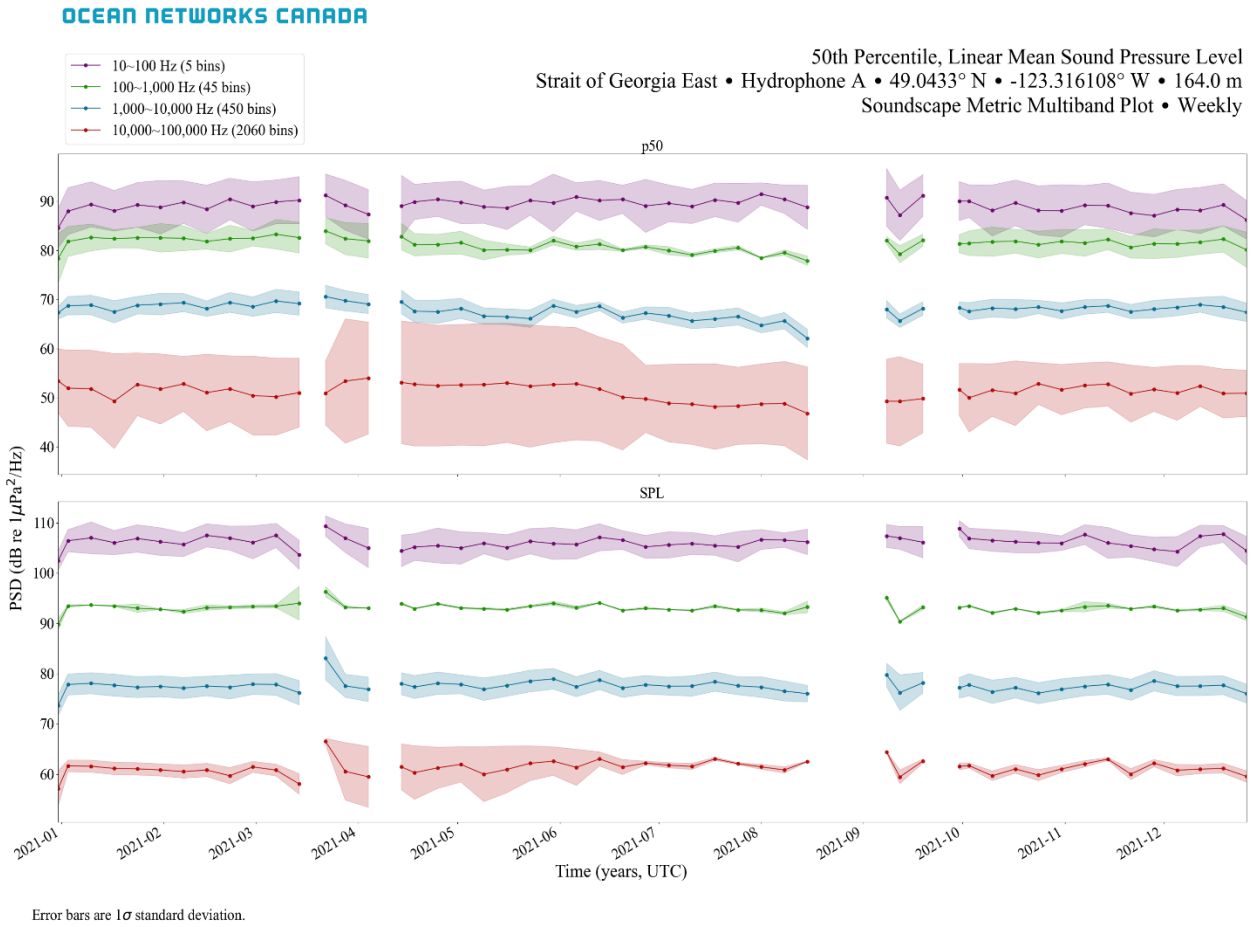

Figure 1. A plot of the soundscape metric multiband data product using the Oceans 3.0 API DP request orderDataProduct[12] to request and download the SPD weekly files this plot is based off of.