

Problem Statement and Goals

Software Engineering

Team 21, Alkalytics
Sumanya Gulati
Kate Min
Jennifer Ye
Jason Tran

Table 1: Revision History

Date	Developer(s)	Change
09-19-2024	Jennifer Ye	Initial Rough Work
09-23-2024	Jennifer Ye	Final Edits After Team Meet
02-22-2025	Jennifer Ye, Kate Min	Incorporated Feedback

1 Problem Statement

This project aims to aid in the data management and analysis of an ocean alkalinity enhancement experiment process.

1.1 Background

This project aims to aid in the data management and analysis of an ocean alkalinity enhancement experiment process. Rising global temperatures are impacting the pH balance of the oceans, which in turn affects their natural ability to absorb carbon dioxide (CO_2) from the atmosphere. The research is working towards a scalable process to capture CO_2 emissions using Bipolar Membrane Electrodialysis (BMED) technology. This process separates a salt solution into distinct acid and base components, with the base solution meant to be returned to the ocean.

The goal of the experimental study is to increase the ocean's pH levels to enhance its ability to absorb more CO_2 , helping to reduce its warming effect and mitigate global temperature rise. Currently, the production process for generating this base is on a small scale and is still being refined. A significant number of experiments and data collection would be required to scale it up globally,

necessitating a more expansive production operation. Optimization of the experimental data is critical to improve process efficiency. This presents a big data software challenge, requiring the ability to identify and fine-tune specific parameters to achieve ideal conditions.

1.2 Problem

Currently, all experiment data is stored in separate Excel spreadsheets, with one file for each day's worth of experiments. This method of data collection and storage is becoming increasingly inefficient and unsustainable as the study grows. The research team is forced to manually sift through the spreadsheets to be able to plot and analyze any results. In some instances, they must also modify certain data values, which can only be done locating the correct file and searching through it. This is an extremely tedious process, which will only become more prevalent with time as the experiments continue to run. The team aims to develop a solution to efficiently store and manage the large volume of data being collected. This solution must ensure that all data is properly labelled, organized, searchable, and have the ability to compare datasets for efficient analysis.

1.3 Inputs and Outputs

The data is being stored in Excel spreadsheets. However, the data can be exported to a Comma Separated Values (CSV) file. This will be the main input for the Alkalytics project. The BMED device automatically exports the collected data as a CSV file. There are no other input file types.

There are two possible outputs for this project, one of which could be an exportable CSV file of the desired data, or a visual plot of the desired data. The choice of output will depend on the user and their specific needs.

1.4 Stakeholders

The ongoing experimental study is currently of small scale.

The main stakeholders currently are:

- **Dr. Charles de Lannoy:** The project supervisor.
Dr. de Lannoy is a chemical engineering professor at McMaster University and leads the *de Lannoy Lab*, the research conducting the ocean alkalinity enhancement experiments.
- **Bassel Abdelkader:** The secondary supervisor.
A Postdoctoral researcher working alongside Dr. de Lannoy to run the experiments and manage the data.
- **Current students/members of the lab**
Student researchers involved in the experimental study may participate for

a limited time. Although they still interact with the experimental data, they may not be considered a primary stakeholder.

- **Lab funding sources**

Organizations or individuals funding the research project. They may be interested in new tools being developed to enhance the data management and analysis.

1.5 Environment

This software will be compatible with any browser on a computer that is connected to the internet. This software will also be compatible with Windows 10 and 11. There is no hardware component.

2 Goals

This project has a total of six goals.

2.1 Consolidate Organized Accurate Data with Proper Labels

The raw data input for this project is a CSV file exported from Excel. With hundreds of experiments, each containing ten or more attributes, a single dataset can contain over thousands of recorded values. As the study expands, the current data format becomes unmanageable. Therefore, one of the primary goals is to migrate all the existing data to a solution that is both scalable and extendable. This migration must ensure that the data is properly labelled according to its original structure. This goal can be measured by the percentage of data successfully imported and accurately labelled.

2.2 Querying from Multiple Sets

Information about experiments is stored in a single spreadsheet, with each record having its own dedicated raw data file. Although the data is organized within these files, it is difficult to find and query multiple datasets efficiently. The only method the research team currently has to achieve this is manually finding the spreadsheet filename with the corresponding experiment date. This project should be able to pull data across sheets quickly and efficiently. This goal can be measured by the time it takes for a query to complete.

2.3 Inter-Parameter Comparability

To improve the data analysis, the product should allow for comparisons of multiple parameters. This could include a simple comparison between two or more attributes, or narrowing down the search window. One of the goals for this project is to provide users with the control over the data they wish to view, and

the ability to define the specific ranges for that data. This goal can be measured by the number of parameters that can be searched for/compared at once.

2.4 Web Interface

The product should have a web interface to enable the user to interact with the backend processes easily. This interface should be user-friendly and display all necessary functions for data input and retrieval. The returned data should be presented in a readable format. This goal can be measured by how many of the required tasks it can successfully perform, and how intuitive it is for the user.

2.5 Flexibility in Expansion (i.e. adding additional parameters)

Although one of the core goals of this project is to migrate the existing data to a better solution, the product should be able to support future expansion. As the experiment evolves, the research team may identify new parameters that may need to be recorded, or they may want this tool to accomodate other experiments that do not follow the same format. The product should be adaptable to handle such new changes and serve as a general data recording tool. This goal can be measured by How well the product handles unconventional data.

2.6 Visual Plot of Specified Parameters and Results

The product should be capable of generating visual representations of the retrieved data. Users should be allowed to specify the desired format for these visualizations, which should also ideally be exportable. The data must be presented in a way that is clear and easy to understand for it to be meaningful to the user. This goal can be measured by the quality of the generated visualizations.

3 Stretch Goals

Stretch goals show what additional features that could be implemented once the primary goals of the project are achieved. There is a total of four stretch goals for this project.

3.1 Automatically Download Data from Data Collection Device

The current experiment setup uses a BMED device that collects the raw experiment data. This device only stores a certain number of experiment trials, and the data must also be manually exported daily. If the data is not exported within a week, it will be overwritten. As a stretch goal, the product could be integrated with this device to automatically upload the data either after each experiment or when a predetermined limit is reached. This is a stretch goal as

it requires a deeper understanding from the developing team of how the device records the data. This goal can be measured by the frequency of successful automatic downloads and by the amount of data retained without loss.

3.2 Dynamic Dashboard to Generate Comparison Reports

As an extension of one of the main project goals, the visual plots can be enhanced with a dynamic dashboard that includes a brief written summary of the requested data requested along with the plotted graph. Users could also be given the ability to select the type of visual plot. The dashboard could dynamically adapt to additional information, user interactions and removals without having to regenerate the entire plot. This goal can be measured by how well the plots change dynamically.

3.3 Mobile Development and Accessibility

This project is intended to have a website interface. As a stretch goal, the project could be extended to include a mobile app for improved ease of access and accessibility. Both the website and mobile app could incorporate more specific accessibility features, such as high-contrast colour schemes, screen reader compatibility, larger fonts and responsiveness for users with visual impairments. This goal can be measured by how many of these extra features are successfully implemented.

3.4 Machine Learning Analysis and Projections

One major stretch goal is to integrate artificial intelligence into the data analysis. By incorporating a machine learning algorithm, the system could automatically query data, customise queries, analyze the data, and generate a customizable dashboard based on user prompts. The machine learning component could also be trained to provide predictions based on previous analyses or specific user prompts.

4 Challenge Level and Extras

The assigned challenge level for this project is general. This project does not have a direct need for extended research on the development side, however, features and abilities required to develop the product may have its own challenges. There are two extra components the team has chosen to complete for this project.

- Usability Testing
- User documentation

These extras may be subject to change. However, the team has determined that they will help to ensure the project progresses smoothly and keep stakeholders informed about its status, development, and usage.

Appendix — Reflection

1. What went well while writing this deliverable?

When writing this deliverable, it was simple in a sense where many of the topics and ideas presented here were talked about before hand with both the team and the supervisors. It was more of a matter of expanding on ideas and topics in further detail to ensure that the problem and goals were aligned with previous conversations. This allowed the writing process to be more time consuming than actually being difficult.

2. What pain points did you experience during this deliverable, and how did you resolve them?

The biggest pain point during this deliverable was figuring out how to properly explain the problem and goals. Many if not all ideas presented were simply ideas. It was quite difficult to be able to not only explain the goals but to also connect it back to why they are important to have in this project. Displaying the understanding of how pieces fit together and explaining why certain things are in scope was the most challenging part of this deliverable.

3. How did you and your team adjust the scope of your goals to ensure they are suitable for a Capstone project (not overly ambitious but also of appropriate complexity for a senior design project)?

As a team, this was very much a concern in the beginning. This resulted in our team setting up strong measures of communication. Before even starting on this deliverable we made sure to go through each section with the supervisor. We had discussions throughout the meeting where the team brings a more technical point of view to aid in the supervisor's needs and wants. It was through this process that we did not have to adjust the scope of the groups goals much as we started with the a strong foundation and understanding with all parties involved.