

Forecasting Ocean Acidification Rubric

DS 4002 – Emilie Yang

Due: May 32, 2025

Submission format: GitHub repo and other files on Canvas

Individual Assignment

General Description: Use this rubric to work through your project requirements. Submit a link to your Github repository.

Why am I doing this? Time series analysis is an essential aspect of statistical analysis that allows you to understand trends, patterns, and forecast data. This case study will give you the foundations for using time-series analysis in a real-world issue: understanding the causes and effects of ocean acidification.

What am I going to do? You will be choosing a time-series model analyzing trends of oceanic pH and forecast future pH values based on exogenous variables. First, you will prepare the given data by cleaning the datasets. Then, you'll perform an exploratory data analysis (EDA) to visualize patterns in data. Finally, you will perform your analysis and use performance metrics to measure the accuracy of the model.

Use the given GitHub repository (https://github.com/ettyang/DSproject_CS3) to download the data sets.

Analysis tasks:

- Perform EDA to find patterns and correlations between each variable
- Forecast future pH values based on historical data (CO2 levels, sea surface temperature, global temperature, volcanic eruptions)
- Forecast future plankton populations based on historical pH data

Deliverables:

- Written summary: give context to the project and discuss results
- GitHub repository: include code, data, and outputs

Tips for success:

- Look at previous research on your chosen model
- Talk to classmates and instructors for guidance and feedback
- Be clear in your process- don't be overwhelmed by all the steps! Create an outline if you need
- Name all folders and files in your GitHub clearly so that people can easily follow along

How will I know I have succeeded? You will meet expectations on this case study when you follow the criteria in the rubric below.

Formatting	<ul style="list-style-type: none">• One page summary:<ul style="list-style-type: none">○ Submit a PDF file
------------	--

	<ul style="list-style-type: none"> ● GitHub repository folders (submit link): <ul style="list-style-type: none"> ○ README.md ○ LICENSE.md ○ SCRIPTS folder ○ DATA folder ○ OUTPUT folder ○ REFERENCES.md
Written summary	<ul style="list-style-type: none"> ● <u>Goal</u>: present the importance/context of the project, briefly discuss process and results to engage with potential users ● Be brief (1-2 pages)- this isn't meant to be a scientific paper ● Include figures when needed ● Components: <ul style="list-style-type: none"> ○ Title, name, date, course title ○ Context of the project ○ Why is this project important? ○ Analysis process ○ Results/discussion ○ References
GitHub README.md	<ul style="list-style-type: none"> ● <u>Goal</u>: summarize your process for other people to follow ● This should be concise, easy to follow, and refer to repo folders/files by name ● Components; <ul style="list-style-type: none"> ○ Software and packages used ○ Outline of repository folders/files ○ Instructions for result reproduction
GitHub LICENSE.md	<ul style="list-style-type: none"> ● <u>Goal</u>: provide citation of your repository when others use your work ● MIT license should be appropriate
GitHub SCRIPTS folder	<ul style="list-style-type: none"> ● <u>Goal</u>: provide source code to reproduce your results <ul style="list-style-type: none"> ○ If you have multiple scripts, make it obvious the order to run the files ○ Include MANY headings and comments in code to show users where they are in the process
GitHub DATA folder	<ul style="list-style-type: none"> ● <u>Goal</u>: provide data for users to use ● Give the folders/files easy names to refer to ● Components: <ul style="list-style-type: none"> ○ Raw and clean datasets ○ Data dictionary (variable name, datatype, potential example)
GitHub OUTPUT folder	<ul style="list-style-type: none"> ● <u>Goal</u>: show users what will be generated in your project <ul style="list-style-type: none"> ○ Include figures from EDA, analysis, etc
GitHub References	<ul style="list-style-type: none"> ● Provide all references used in the project that you found on your own <ul style="list-style-type: none"> ○ IEEE citation style

Acknowledgements: Thanks to Professor Rasero for the rubric format.