Computational Modeling of Phytoplankton Dynamics with Climatic and Ecological Ramifications

Abhinav K. Sharma

Massachusetts Academy of Math and Science

STEM Project

Instructor: Kevin Crowthers, Ph.D.

Table of Contents:

GLP Record Keeping Contract

I, Abhinav K. Sharma, commit to record keeping in accordance with Good Laboratory Practices.

- My experiments and records will be reproducible, traceable, and reliable.
- I will NOT write my notes on scraps of paper, post-it notes, or other disposable items. My notes will go directly into my laboratory notebook.
- My data will be recorded in real-time. If I cannot record data in real-time, I will record raw data as soon as physically possible.
- I will record both qualitative and quantitative observations in my laboratory notebook and laboratory reports.
- My laboratory notebook will include information on the materials and instruments utilized during experimentation.
- I will initial and date over the edge of any material that is taped into my laboratory notebook.
- I will provide a real-time record of any analysis I perform.
- I will use blue or black pen to make entries in my laboratory notebook. I will NOT use pencil.
- I will define ALL abbreviations.
- If I make a mistake in my laboratory notebook, laboratory worksheets, or other written material, I will not obliterate or obscure the mistake. Instead, I will cross out the mistake using a single line. Any empty spaces in tables or partially used notebook pages will be crossed out using a single diagonal line.
- If I record information online (ex. In Google Drive), I will login so that my contributions are traceable.
- I will initial and date each page in my notebook and the front of each laboratory report.

Abhinav Krishan Sharma
Printed name

Abhinav K. Sharma

Signature

19 August 2023
Date

A more detailed description of GLP is located here:
https://docs.google.com/document/d/1zeYoNSniKTc7MlBgTG1SEnhJiCK3UimCvTcKPQcyHGw/edit?usp=sharing

Logbook Etiquette Date: 19 August 2023

For research and engineering purposes, a logbook is considered a legal document and will help in providing documentation for the origin of ideas.

1- When adding something written in Pen- Blue or Black ~~not a Pencil~~ (and DO NOT USE WHITEOUT- ~~mistakes~~ can be corrected by adding the information above the crossed out material and adding your initials and date
2- Don’t worry about neatness- it is a living document but **should be legible but understandable**

3- Page Numbers should be consecutive and located on the top corner of the page- outer edge
4- Do not remove pages
5- Put a line through empty space
6- Neat handwriting
7- **Make an entry every time you work on your project**
8- Make sure your entries are verified by a mentor/ teacher signature and your signature
9- Organize your Notebook: Format

A: Table of Contents
B: Brainstorming and Topic Ideas
C: Project Introduction: Topic, Phrase 1(Testable Question/Engineering Need/Mathematical Conjecture), Phrase 2 + Timeline
D: Communications (i.e. to corresponding authors, mentors, and expert consultation, etc)
E: Draft of Materials and Methods (this can be performed for daily entries if variations occur over the course of the project).
F: Background- ie. competitor/market analysis, criteria/constraints
G: Daily Entries (every time you complete work on the project)

1: Title and Date
2: Short Introduction (putting the experiment/observations into context/objectives)
3: Methods/Materials (if not included in the beginning of the notebook)
Materials become important when someone needs to repeat your experiments
4: Observations/Experimental Data (both RAW and ANALYZED)-

A: graphs/figures
B: data tables
C: pictures
D: sketches or proof of concepts and prototypes (with labels)
E: Decision matrices
E: Ethical responsibility

5: Calculations and Data Analysis (STATISTICS)
6: Final Concluding Remarks

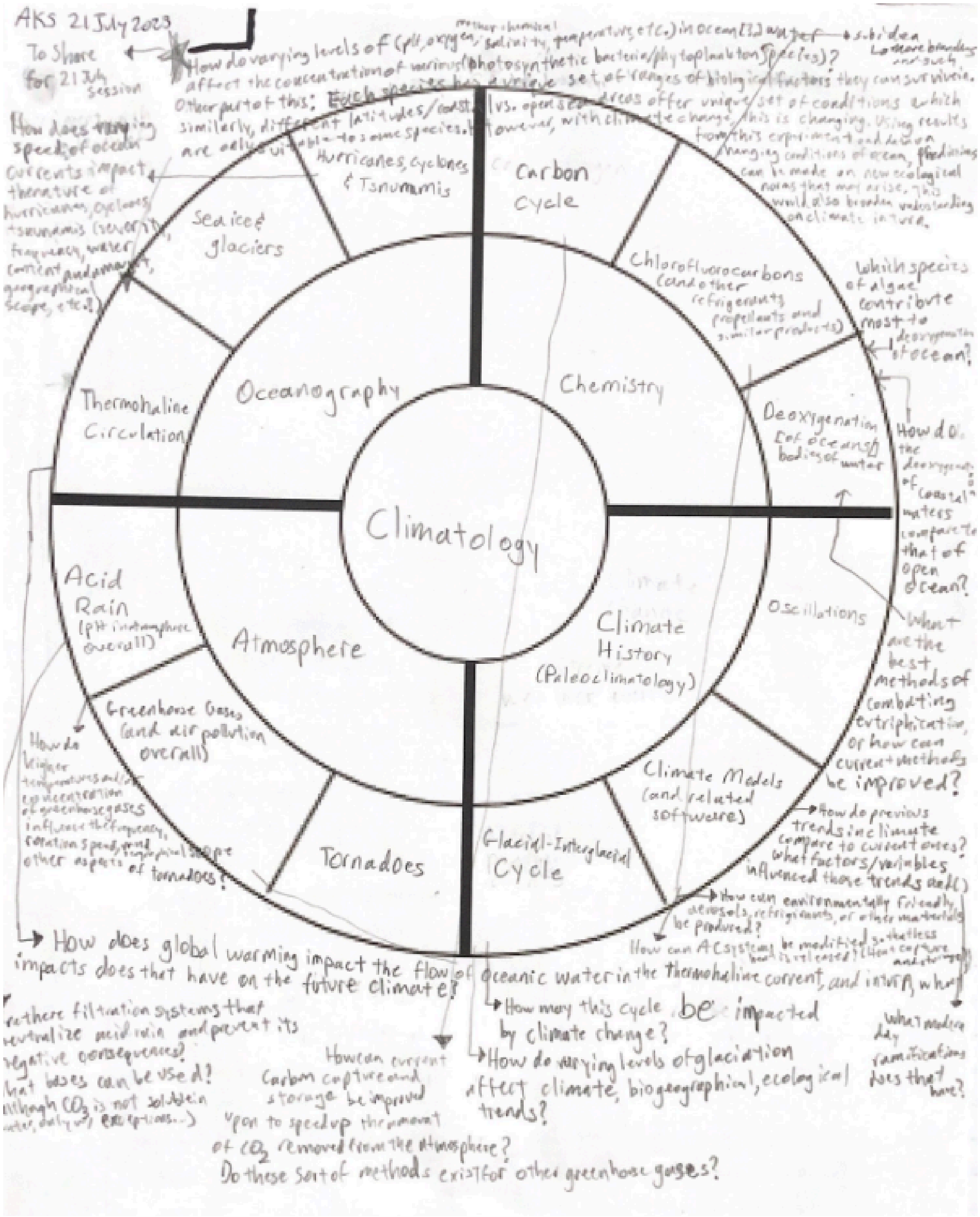
Things to keep in mind:

-You don’t want to have too much blank space
-If you are adding a pre-printed graph or sketch, paste in and sign + date.

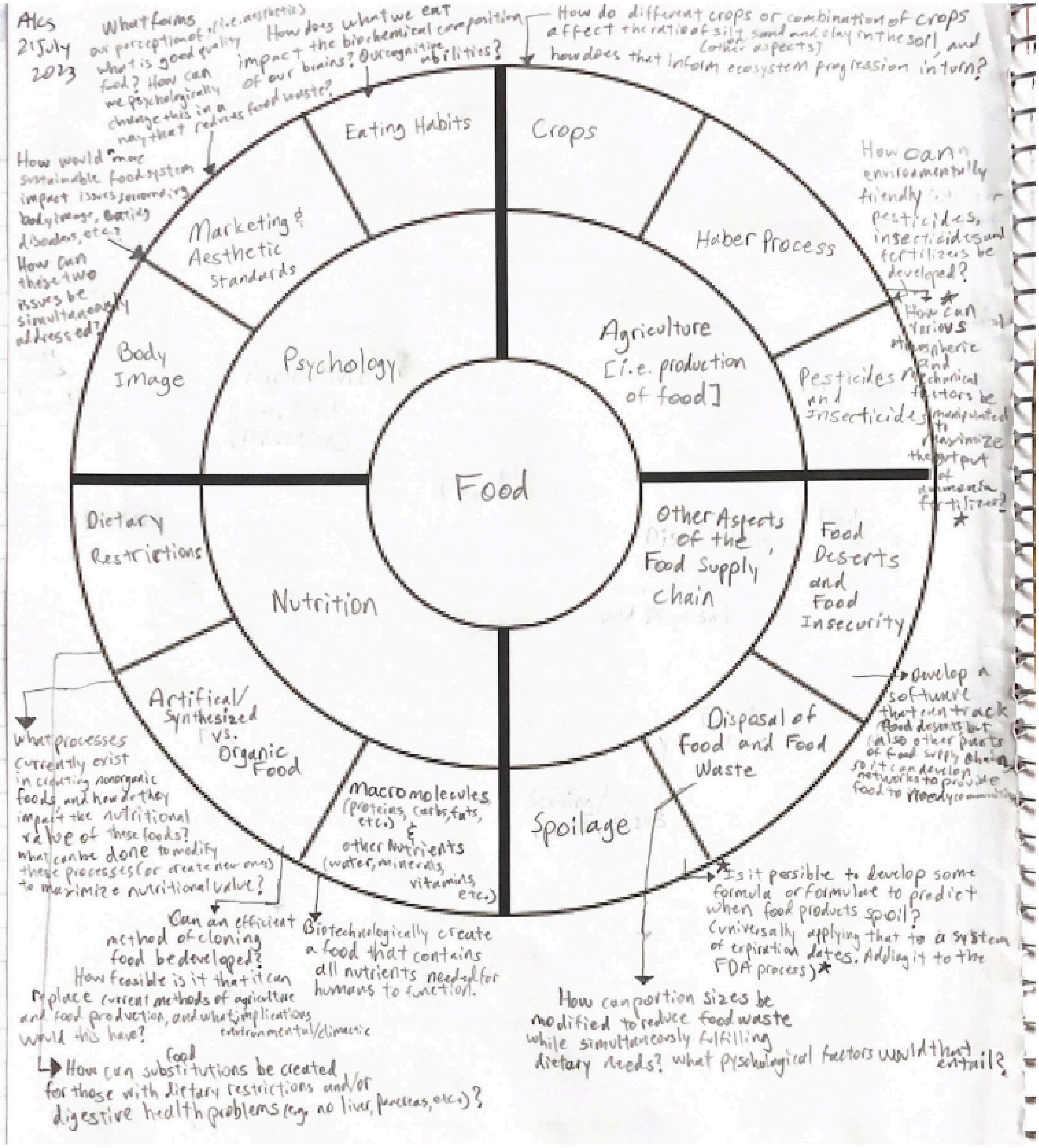
Brainstorming

Pie Diagrams:

Abhinav K. Sharma 4 September 2023 2:37 PM


AKS 21 July 2023
To Share for 21th session
How do varying levels of (pH, oxygen, other chemical, salinity, temperature, etc.) in ocean [?] water affect the concentration of various photosynthetic bacteria/phytoplankton species? Each species has a unique set of ranges of biological factors they can survive in. Other part of this: coastal vs. open sea areas offer unique set of conditions which similarly, different latitudes/coastal are only suitable to some species. However, with climate change, this is changing. Using results from this experiment and data on changing conditions of ocean, predictions can be made on new ecological norms that may arise. This would also broaden understanding on climate in future.
sub-idea
How does varying speed of ocean currents impact the nature of hurricanes, cyclones + tsunamis (severity, frequency, water content and geographical scope, etc.)
Climatology
Oceanography
Chemistry
Atmosphere
Climate History (Paleoclimatology)
Hurricanes, cyclones & Tsunamis
Sea ice & glaciers
Thermohaline Circulation
Carbon Cycle
Chlorofluorocarbons (and other refrigerants propellants and similar products)
Deoxygenation (of oceans & bodies of water)
Oscillations
Climate Models (and related software)
Glacial-Interglacial Cycle
Tornadoes
Greenhouse Gases (and air pollution overall)
Acid Rain (pH in atmosphere overall)
Which species of algae contribute most to deoxygenation of ocean?
How does the deoxygenation of coastal waters compare to that of open ocean?
What are the best methods of combating eutrophication, or how can current methods be improved?
How do previous trends in climate compare to current ones? What factors/variables influenced those trends and
How can environmentally friendly aerosols, refrigerants, or other materials be produced?
How can AC systems be modified so that less heat is released? How capture and
What modern day ramifications does that have?
How do higher temperatures and/or concentration of greenhouse gases influence the frequency, rotation speed, and geographical scope other aspects of tornadoes?
How does global warming impact the flow of oceanic water in the thermohaline current, and in turn, what impacts does that have on the future climate?
How may this cycle be impacted by climate change?
How do varying levels of glaciation affect climate, biogeographical, ecological trends?
How can current carbon capture and storage be improved upon to speed up the amount of CO2 removed from the atmosphere? Do these sort of methods exist for other greenhouse gases?
Are there filtration systems that neutralize acid rain and prevent its negative consequences? What bases can be used? (although CO2 is not soluble in water, only w/ exceptions...)

A major area of interest when first developing a project idea was climatology. Indeed, this field has been a career aspiration of mine. As such, I wanted to do some idea with the climate field. Multiple sectors of the climate system were considered when developing ideas, including atmospheric and oceanic conditions, extreme weather events and historical modeling. Through connecting to my other major idea, ecology, the main idea for this project, i.e. phytoplankton, was developed.


Alcs
21 July
2023
Food
Psychology
Eating Habits
Marketing & Aesthetic Standards
Body Image
Agriculture [i.e. production of food]
Crops
Haber Process
Pesticides and Insecticides
Nutrition
Dietary Restrictions
Artifical/ Synthesized vs. Organic Food
Macromolecules (proteins, carbs, fats, etc.) & other Nutrients (water, minerals, vitamins, etc.)
Other Aspects of the Food Supply chain
Food Deserts and Food Insecurity
Disposal of Food and Food Waste
Spoilage

What forms (i.e. aesthetics) our perception of what is good quality food? How can we psychologically change this in a way that reduces food waste?

How does what we eat impact the biochemical composition of our brains? Our cognitive abilities?

How do different crops or combination of crops affect the ratio of silt, sand and clay in the soil (or other aspects), and how does that inform ecosystem progression in turn?

How would more sustainable food system impact issues concerning body image, eating disorders, etc.?

How can these two issues be simultaneously addressed?

How can environmentally friendly pesticides, insecticides and fertilizers be developed? ★

How can various atmospheric and mechanical factors be manipulated to maximize the output of ammonia fertilizer? ★

What processes currently exist in creating nonorganic foods, and how do they impact the nutritional value of these foods? What can be done to modify these processes (or create new ones) to maximize nutritional value?

Can an efficient method of cloning food be developed? How feasible is it that it can replace current methods of agriculture and food production, and what environmental/climactic implications would this have?

Biotechnologically create a food that contains all nutrients needed for humans to function.

How can food substitutions be created for those with dietary restrictions and/or digestive health problems (e.g. no liver, pancreas, etc.)?

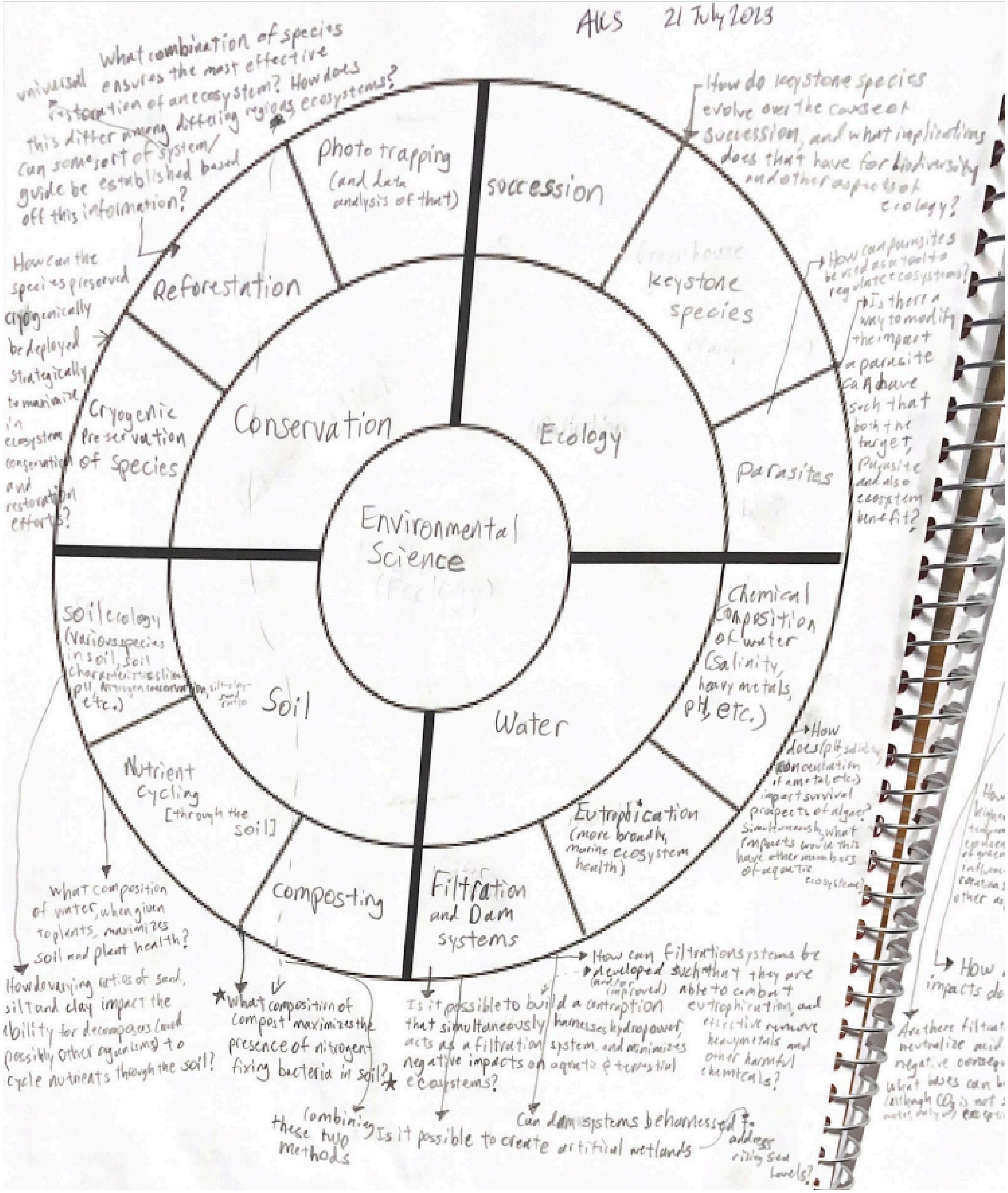
Develop a software that can track food deserts but (also other parts of food supply chain) so it can develop networks to provide food to needy communities

★ Is it possible to develop some formula or formulae to predict when food products spoil? (universally applying that to a system of expiration dates. Adding it to the FDA process) ★

How can portion sizes be modified to reduce food waste while simultaneously fulfilling dietary needs? What pyschological factors would that entail?

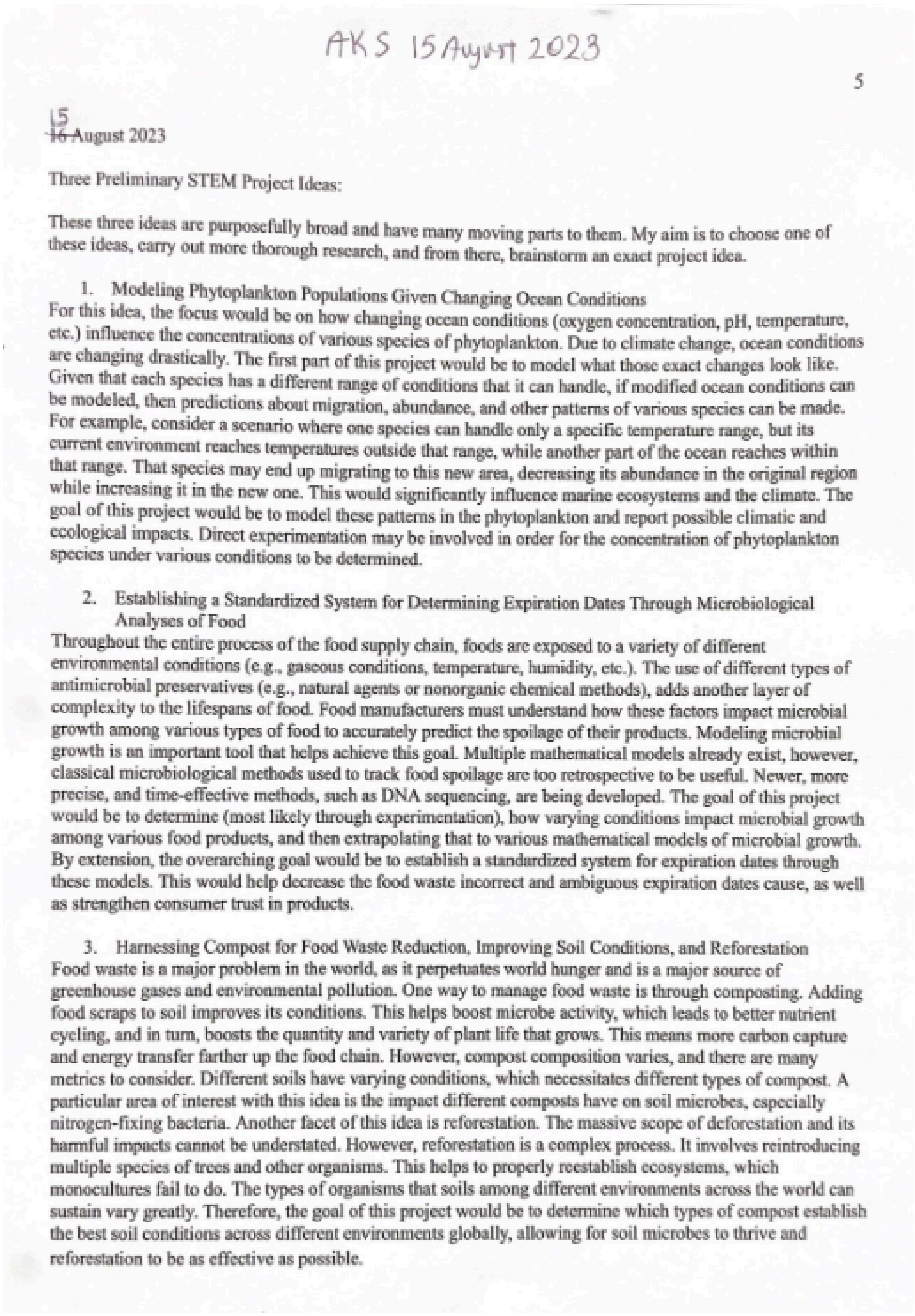
Abhinav K. Sharma
4 September 2023 2:37 PM

Food was considered as a more secondary, back-up category during the brainstorming phase. However, it has been an area of interest nonetheless. Many connections to other fields of science were made based off this simple basic need, including biotechnology, the food supply chain, environmental and agricultural sciences and psychology. A major backup idea that was identified was mathematically modeling microbial growth in different foods as a way of developing a standardized system for creating expiration dates as a means to address food waste.


AKS 21 July 2023
Environmental Science
Conservation
Ecology
Soil
Water
Photo trapping (and data analysis of that)
Succession
Keystone species
Parasites
Reforestation
Cryogenic Preservation of Species
Soil ecology (various species in soil, soil characteristics like pH, nitrogen concentration, silt:clay:sand ratio etc.)
Nutrient Cycling [through the soil]
Composting
Filtration and Dam systems
Eutrophication (more broadly, marine ecosystem health)
Chemical Composition of water (Salinity, heavy metals, pH, etc.)
What combination of species ensures the most effective restoration of an ecosystem? How does this differ among differing regions ecosystems? universal Can some sort of system/guide be established based off this information?
How can the species preserved cryogenically be deployed strategically to maximize in ecosystem conservation and restoration efforts?
How do keystone species evolve over the course of succession, and what implications does that have for biodiversity and other aspects of ecology?
How can parasites be used as a tool to regulate ecosystems?
Is there a way to modify the impact a parasite can have such that both the target, parasite and also ecosystem benefit?
How does (pH, salinity, concentration of a metal, etc.) impact survival prospects of algae? Simultaneously, what impacts would this have other members of aquatic ecosystems?
What composition of water, when given to plants, maximizes soil and plant health?
How do varying ratios of sand, silt and clay impact the ability for decomposers (and possibly other organisms) to cycle nutrients through the soil?
★What composition of compost maximizes the presence of nitrogen fixing bacteria in soil?★
Is it possible to build a contraption that simultaneously harnesses hydropower, acts as a filtration system, and minimizes negative impacts on aquatic & terrestrial ecosystems?
How can filtration systems be developed (and/or improved) such that they are able to combat eutrophication, and effective remove heavy metals and other harmful chemicals?
Combining these two methods
Is it possible to create artificial wetlands
Can dam systems be harnessed to address rising sea levels?
How impacts do
Are there filt neutralize acid negative conseq What bases can b (although CO₂ is not water, only w) exceptio
other as

Abhinav K. Sharma
4 September 2023 2:37 PM

As with climatology, environmental science is an area of potential career aspiration. Multiple aspects of environmental science were considered while brainstorming. This included soil, water, conservation, ecology, mitigation strategies like dams, abiotic factors, composting, and cryogenic preservation. A heavier focus was placed more on environmental chemistry and abiotic factors, more on ideas that would involve scientific experimentation.



AKS 15 August 2023

5

~~16~~ 15 August 2023

Three Preliminary STEM Project Ideas:

These three ideas are purposefully broad and have many moving parts to them. My aim is to choose one of these ideas, carry out more thorough research, and from there, brainstorm an exact project idea.

1. Modeling Phytoplankton Populations Given Changing Ocean Conditions

For this idea, the focus would be on how changing ocean conditions (oxygen concentration, pH, temperature, etc.) influence the concentrations of various species of phytoplankton. Due to climate change, ocean conditions are changing drastically. The first part of this project would be to model what those exact changes look like. Given that each species has a different range of conditions that it can handle, if modified ocean conditions can be modeled, then predictions about migration, abundance, and other patterns of various species can be made. For example, consider a scenario where one species can handle only a specific temperature range, but its current environment reaches temperatures outside that range, while another part of the ocean reaches within that range. That species may end up migrating to this new area, decreasing its abundance in the original region while increasing it in the new one. This would significantly influence marine ecosystems and the climate. The goal of this project would be to model these patterns in the phytoplankton and report possible climatic and ecological impacts. Direct experimentation may be involved in order for the concentration of phytoplankton species under various conditions to be determined.

2. Establishing a Standardized System for Determining Expiration Dates Through Microbiological Analyses of Food

Throughout the entire process of the food supply chain, foods are exposed to a variety of different environmental conditions (e.g., gaseous conditions, temperature, humidity, etc.). The use of different types of antimicrobial preservatives (e.g., natural agents or nonorganic chemical methods), adds another layer of complexity to the lifespans of food. Food manufacturers must understand how these factors impact microbial growth among various types of food to accurately predict the spoilage of their products. Modeling microbial growth is an important tool that helps achieve this goal. Multiple mathematical models already exist, however, classical microbiological methods used to track food spoilage are too retrospective to be useful. Newer, more precise, and time-effective methods, such as DNA sequencing, are being developed. The goal of this project would be to determine (most likely through experimentation), how varying conditions impact microbial growth among various food products, and then extrapolating that to various mathematical models of microbial growth. By extension, the overarching goal would be to establish a standardized system for expiration dates through these models. This would help decrease the food waste incorrect and ambiguous expiration dates cause, as well as strengthen consumer trust in products.

3. Harnessing Compost for Food Waste Reduction, Improving Soil Conditions, and Reforestation

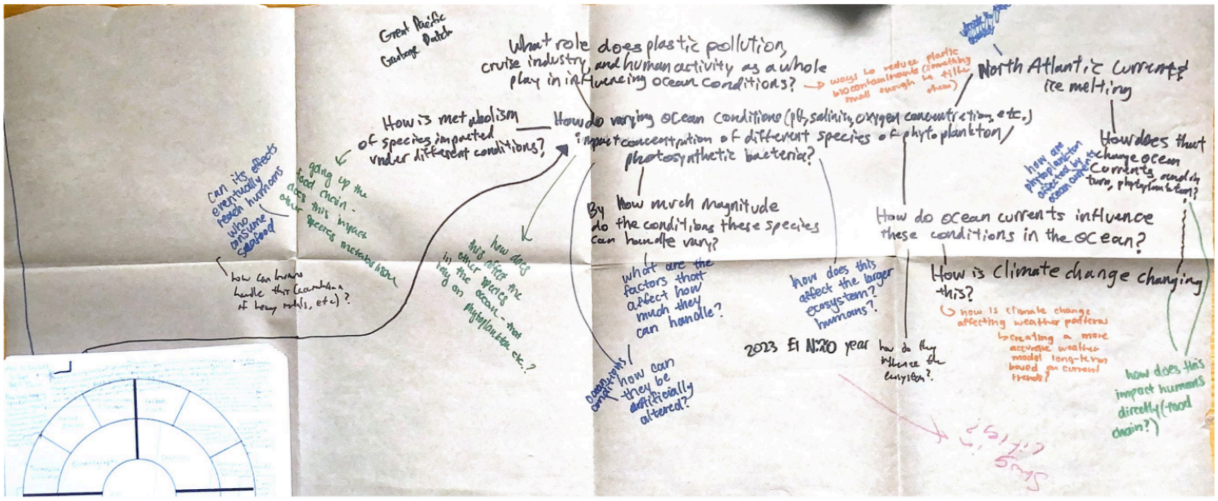
Food waste is a major problem in the world, as it perpetuates world hunger and is a major source of greenhouse gases and environmental pollution. One way to manage food waste is through composting. Adding food scraps to soil improves its conditions. This helps boost microbe activity, which leads to better nutrient cycling, and in turn, boosts the quantity and variety of plant life that grows. This means more carbon capture and energy transfer farther up the food chain. However, compost composition varies, and there are many metrics to consider. Different soils have varying conditions, which necessitates different types of compost. A particular area of interest with this idea is the impact different composts have on soil microbes, especially nitrogen-fixing bacteria. Another facet of this idea is reforestation. The massive scope of deforestation and its harmful impacts cannot be understated. However, reforestation is a complex process. It involves reintroducing multiple species of trees and other organisms. This helps to properly reestablish ecosystems, which monocultures fail to do. The types of organisms that soils among different environments across the world can sustain vary greatly. Therefore, the goal of this project would be to determine which types of compost establish the best soil conditions across different environments globally, allowing for soil microbes to thrive and reforestation to be as effective as possible.

Abhinav K. Sharma

4 September 2023 2:40 PM

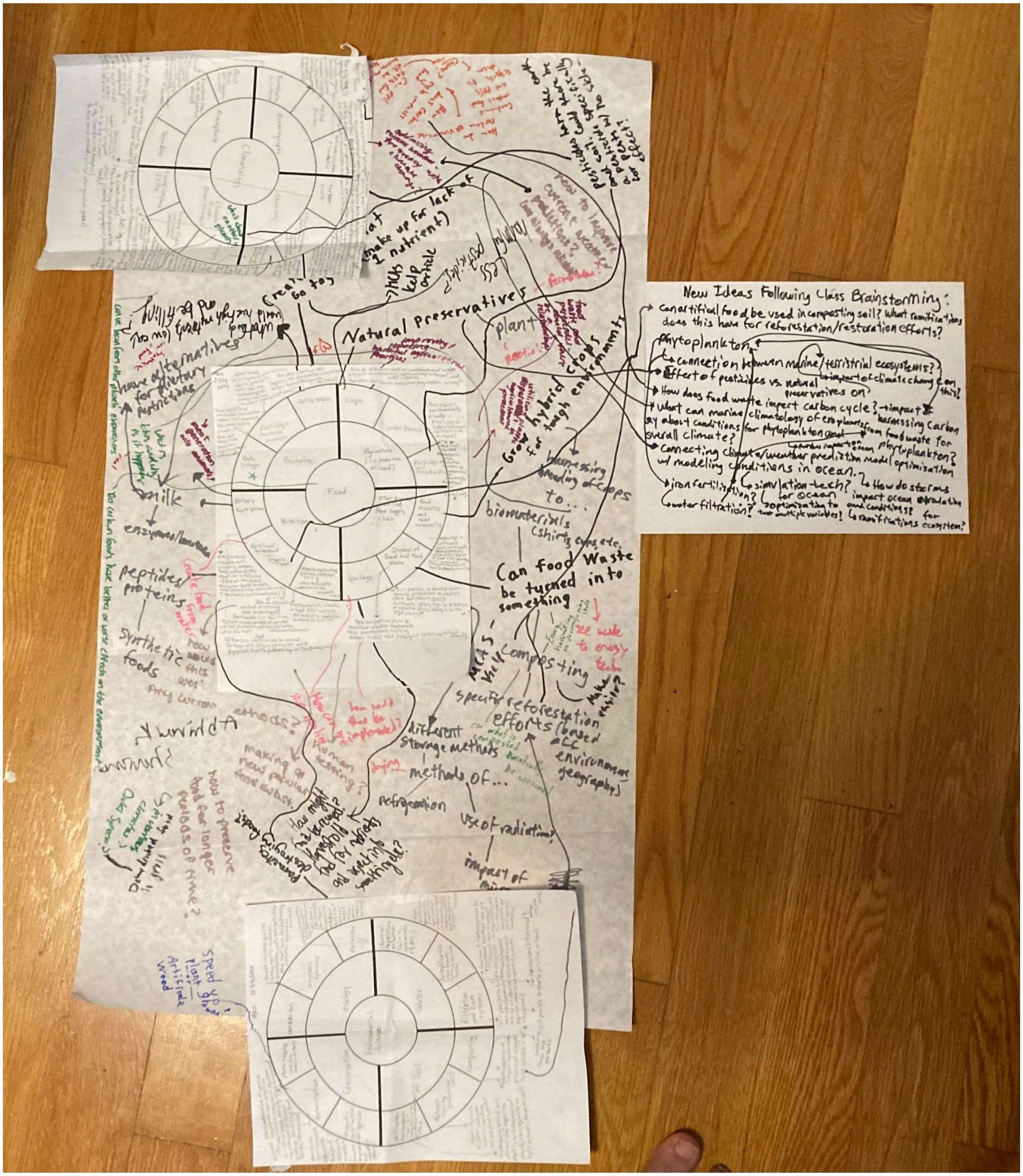
As part of a STEM summer brainstorming assignment, three preliminary ideas were developed. The first idea involved modeling phytoplankton given global warming-induced changes in oceanic conditions. The second idea involved identifying ideal composts for different soils across the world and implementing them as a means to enhance ecological restoration efforts. The third idea involved modeling microbial growth in various food products as a way of developing a standardized formula for expiration dates and thereby reducing food waste that results thereof.

Mindmaps


Great Pacific Garbage Patch
What role does plastic pollution, cruise industry, and human activity as a whole play in influencing ocean conditions?
ways to reduce plastic bio contaminants (something small enough to filter ocean)
North Atlantic currents
ice melting
How is metabolism of species impacted under different conditions?
How do varying ocean conditions (pH, salinity, oxygen concentration, etc.) impact concentration of different species of phytoplankton/ photosynthetic bacteria?
can its effects eventually reach humans who consume seafood
going up the food chain - does this impact other species metabolism
how can humans handle this (accumulation of heavy metals, etc)?
how does this affect the other species in the ocean - that rely on phytoplankton etc.?
By How much magnitude do the conditions these species can handle vary?
what are the factors that affect how much they can handle?
ocean conditions / how can they be artificially altered?
how does this affect the larger ecosystem? humans?
How do ocean currents influence these conditions in the ocean?
How does that change ocean currents and in turn, phytoplankton?
how are phytoplankton affected by ocean currents
How is climate change changing this?
how is climate change affecting weather patterns
creating a more accurate weather model long-term based on current trends?
2023 El Niño year
how do they influence the ecosystem?
how does this impact humans directly (food chain?)

Abhinav K. Sharma
4 September 2023 2:37 PM

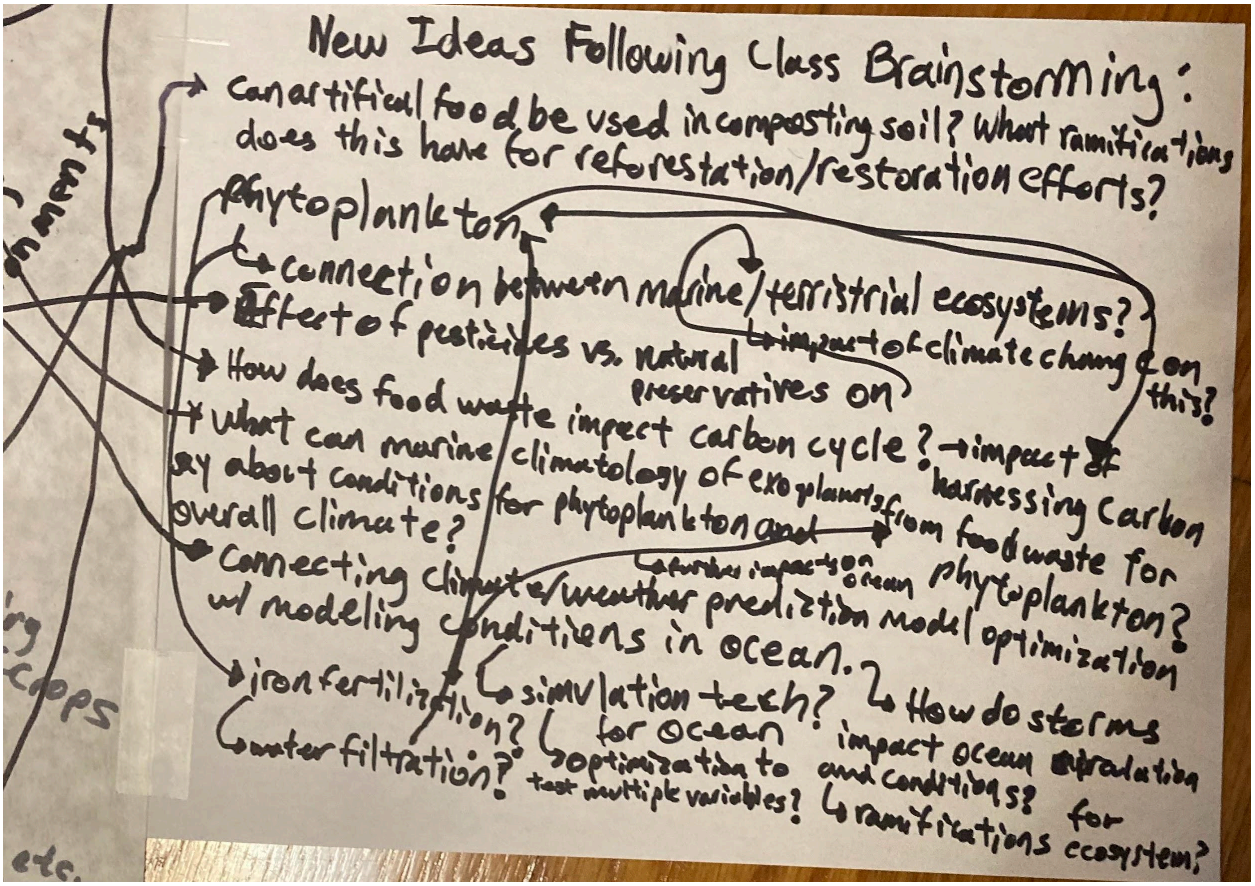
This is the Mindmap generated during the Summer Brainstorming session in July. As opposed to using the pie chart as a starting-off point for brainstorming, the center of the mindmap was an experimental question relating to phytoplankton. While this central idea was very narrow, many new insights were generated from the ideas of others. Their input made me consider more deeply the ecological role of phytoplankton, ocean currents and oceanic conditions, melting sea ice, among other factors.


New Ideas Following Class Brainstorming:
Can artificial food be used in composting soil? What ramifications does this have for reforestation/restoration efforts?
Phytoplankton
connection between marine/terristrial ecosystems?
impact of climate change on this?
Effect of pesticides vs. natural preservatives on
How does food waste impact carbon cycle? → impact of harnessing carbon from food waste for phytoplankton?
What can marine/climatology of exoplanets say about conditions for phytoplankton and overall climate?
Connecting climate/weather prediction model optimization w/ modeling conditions in ocean.
simulation tech? for ocean optimization to test multiple variables?
How do storms impact ocean and conditions? ramifications for ecosystem?
iron fertilization?
water filtration?
Food
Climatology
Natural Preservatives
plant
Grow hybrid crops for tough environments
harnessing breeding of crops to...
make up for lack of 1 nutrient)
kelp
have alternatives for dietary restrictions
milk
enzymes
peptides/proteins
synthetic foods
create food from waste
biomaterials (shirts, cups, etc.
Can food Waste be turned into something
composting
see waste to energy tech
specific reforestation efforts (based off environment geography)
different storage methods
methods of...
refrigeration
drying
use of radiation?
impact of
how to improve current accuracy of predictions?
how to preserve food for longer periods of time?
making a new popular food
Speed up plant growth
Artificial wood

Abhinav K. Sharma

4 September 2023 2:42 PM

This mindmap unites topics that were brainstormed such that all three pie charts were used. For this mindmap, food was at the center of the page. Most of the ideas generated related to addressing food waste via varying fields of science, including agriculture, biotechnology, conservation, among others. Notably, biotechnology was noted as an idea for synthesizing foods to accommodate dietary restrictions. This was the main mind map, the one parents and all other classmates added to.



New Ideas Following Class Brainstorming:

Can artifical food be used in composting soil? What ramifications does this have for reforestation/restoration efforts?

Phytoplankton

connection between marine/territrial ecosystems?

impact of climate change on this?

Effect of pesticides vs. natural preservatives on

How does food waste impact carbon cycle? → impact of harnessing carbon from food waste for phytoplankton?

further impacts on ocean

What can marine/climatology of exoplanets say about conditions for phytoplankton and overall climate?

Connecting climate/weather prediction model optimization w/ modeling conditions in ocean.

simulation tech? for ocean

optimization to test multiple variables?

How do storms impact ocean circulation and conditions?

ramifications for ecosystem?

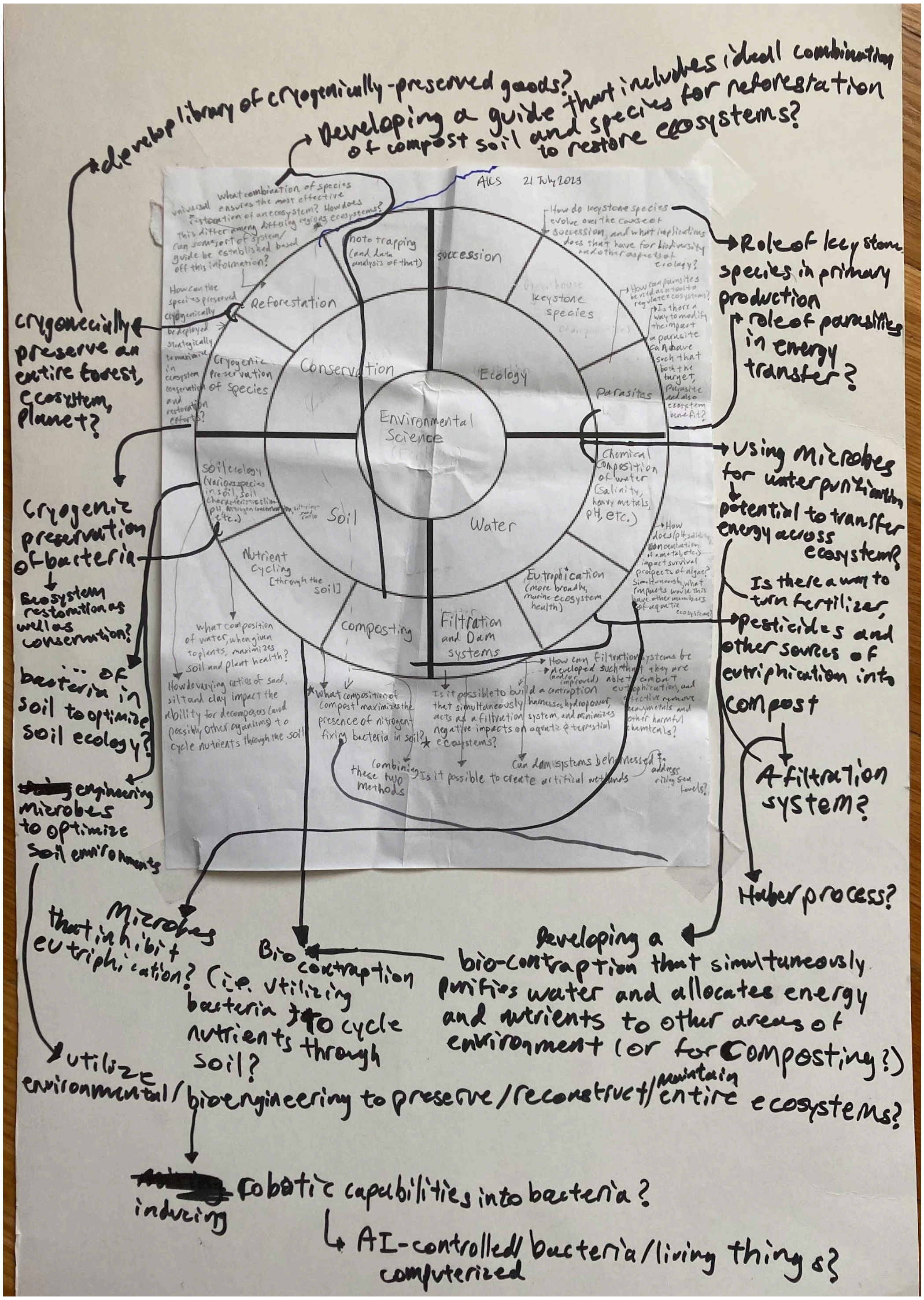
iron fertilization?

water filtration?

Abhinav K. Sharma

4 September 2023 2:42 PM

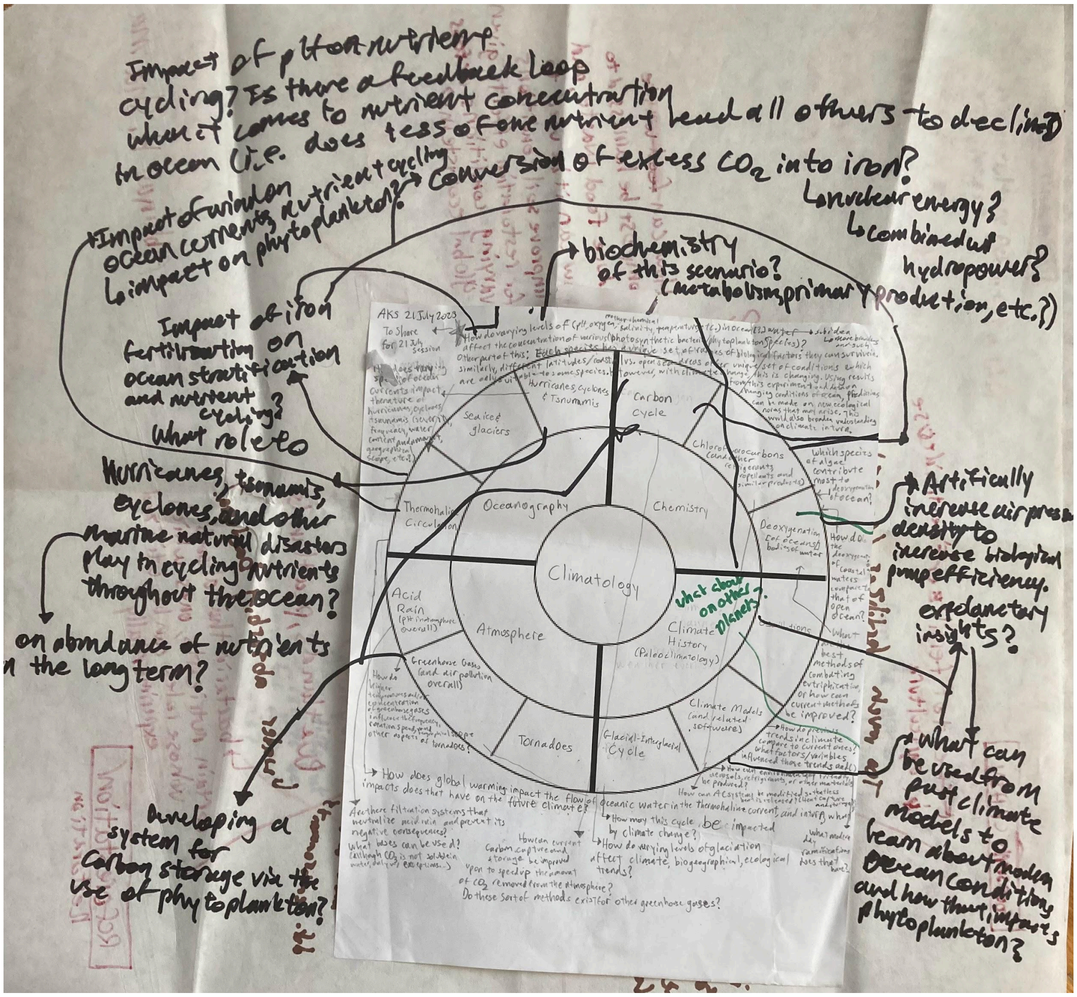
This is a specific part of the mind map pasted above. With the goal being to do a project relating to phytoplankton, an entire section was dedicated to connecting all other ideas from the rest of the large mindmap to any possible project ideas regarding phytoplankton. Ideas range from specific relationships with various environmental parameters to more broad ones, involving large-scale phytoplankton modeling. Which method is chosen for this project (laboratory or Computer science project) is a major source of consideration.


develop library of cryogenically-preserved goods?
Developing a guide that includes ideal combination of compost soil and species for reforestation to restore ecosystems?
AKS 21 July 2023
Environmental Science
Conservation
Ecology
Soil
Water
Reforestation
Photo trapping (and data analysis of that)
Succession
Keystone species
Parasites
Cryogenic Preservation of Species
Chemical Composition of water (Salinity, heavy metals, pH, etc.)
Soil ecology (various species in soil, soil characteristics like pH, nitrogen concentration, silt-clay-sand ratio etc.)
Nutrient cycling [through the soil]
Composting
Filtration and Dam systems
Eutrophication (more broadly, marine ecosystem health)
What combination of species ensures the most effective universal restoration of an ecosystem? How does this differ among differing regions ecosystems? Can some sort of system/guide be established based off this information?
How can the species preserved cryogenically be deployed strategically to maximize in ecosystem conservation and restoration efforts?
How do keystone species evolve over the course of succession, and what implications does that have for biodiversity and other aspects of ecology?
How can parasites be used as a tool to regulate ecosystems?
Is there a way to modify the impact a parasite can have such that both the target, parasite and also ecosystem benefit?
How does (pH, salinity, concentration of a metal, etc.) impact survival prospects of algae? Simultaneously, what impacts would this have other members of aquatic ecosystems?
What composition of water, when given to plants, maximizes soil and plant health?
How do varying ratios of sand, silt and clay impact the ability for decomposers (and possibly other organisms) to cycle nutrients through the soil
What composition of compost maximizes the presence of nitrogen fixing bacteria in soil?
Is it possible to build a contraption that simultaneously harnesses hydropower, acts as a filtration system, and minimizes negative impacts on aquatic & terrestial ecosystems?
How can filtration systems be developed (and/or improved) such that they are able to combat eutrophication, and effective remove heavy metals and other harmful chemicals?
combining these two methods
Is it possible to create artifical wetlands
Can dam systems be harnessed to address rising sea levels?
cryogenically preserve an entire forest, ecosystem, planet?
Role of key stone species in primary production
role of parasites in energy transfer?
Using microbes for water purification
potential to transfer energy across ecosystems
cryogenic preservation of bacteria
Ecosystem restoration as well as conservation?
... of bacteria in soil to optimize soil ecology?
Is there a way to turn fertilizer, pesticides and other sources of eutriphication into compost
A filtration system?
Haber process?
~~[illegible]~~ engineering microbes to optimize soil environments
Microbes that inhibit eutriphication?
Bio contraption (i.e. utilizing bacteria to cycle nutrients through soil?
Developing a bio-contraption that simultuneously purifies water and allocates energy and nutrients to other areas of environment (or for Composting?)
utilize environmental/bioengineering to preserve/reconstruct/maintain entire ecosystems?
~~[illegible]~~ inducing robotic capabilities into bacteria?
AI-controlled/computerized bacteria/living things?

Abhinav K. Sharma

4 September 2023 4:43 PM

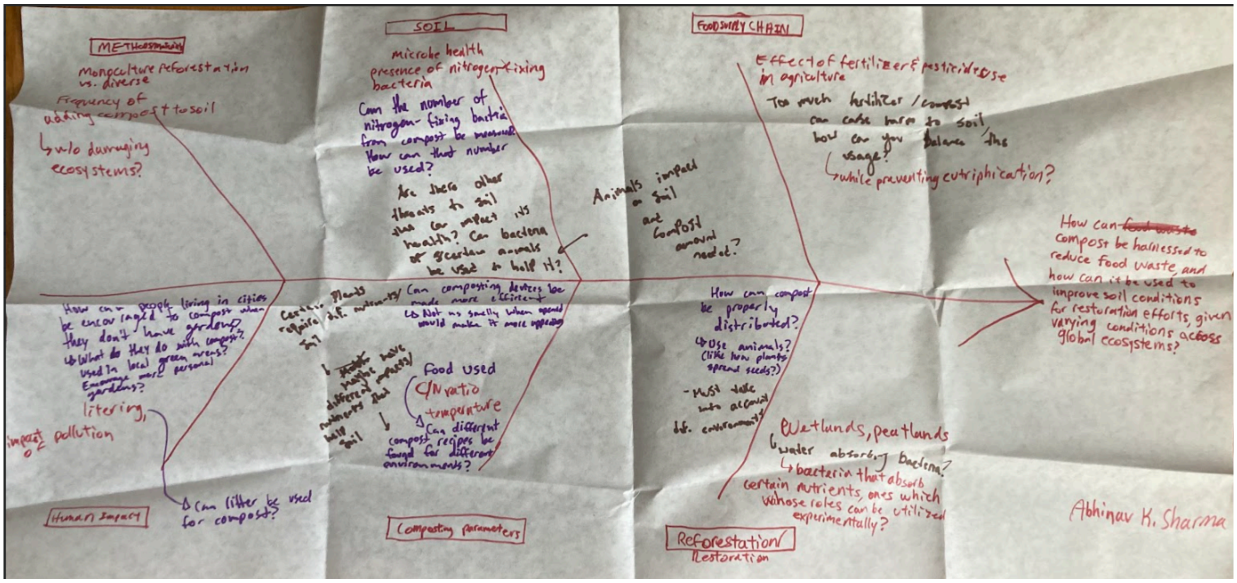
This is the mind-map I created for my environmental science pie chart. For each of the quadrants, ideas are expanded upon, and outside fields are related. For example, photo trapping is expanded to bio contraptions, and even inducing robotic capabilities into bacteria, and from there, AI-controlled living things. There is also a heavy focus placed on cryogenic preservation and soil science. The most appealing idea is modeling ideal composts for different regions of the world as a means of maximizing ecological restoration efforts.


Impact of pH on nutrient cycling? Is there a feedback loop when it comes to nutrient concentration in ocean (i.e. does less of one nutrient lead all others to decline?)
Conversion of excess CO_2 into iron?
Impact of wind on ocean currents nutrient cycling
↳ impact on phytoplankton?
↳ nuclear energy?
↳ combined w/ hydropower?
biochemistry of this scenario? (metabolism, primary production, etc.?)
Impact of iron fertilization on ocean stratification & and nutrient cycling?
What role do Hurricanes, tsunamis, cyclones, and other marine natural disasters play in cycling nutrients throughout the ocean?
on abundance of nutrients in the long term?
Artificially increase air pressure density to increase biological pump efficiency.
exoplanetary insights?
What can be used from past climate models to learn about modern ocean conditions and how that impacts phytoplankton?
Developing a system for carbon storage via the use of phytoplankton?
what about on other planets?
AKS 21 July 2023
To Share for 21.24 Session
Climatology
Oceanography
Chemistry
Atmosphere
Climate History (Paleoclimatology)
Sea ice & glaciers
Hurricanes, cyclones & Tsunamis
Carbon Cycle
Chlorofluorocarbons
Deoxygenation
Thermohaline Circulation
Acid Rain (pH in atmosphere overall)
Greenhouse Gases (and air pollution overall)
Tornadoes
Glacial-Interglacial Cycle
Climate Models (and related software)
How does global warming impact the flow of oceanic water in the thermohaline current, and in turn, what impacts does that have on the future climate?
How may this cycle be impacted by climate change?
How do varying levels of glaciation affect climate, biogeographical, ecological trends?
Which species of algae contribute most to deoxygenation?
Are there filtration systems that neutralize acid rain and prevent its negative consequences?
Do these sorts of methods exist for other greenhouse gases?

Abhinav K. Sharma 4 September 2023 4:43 PM

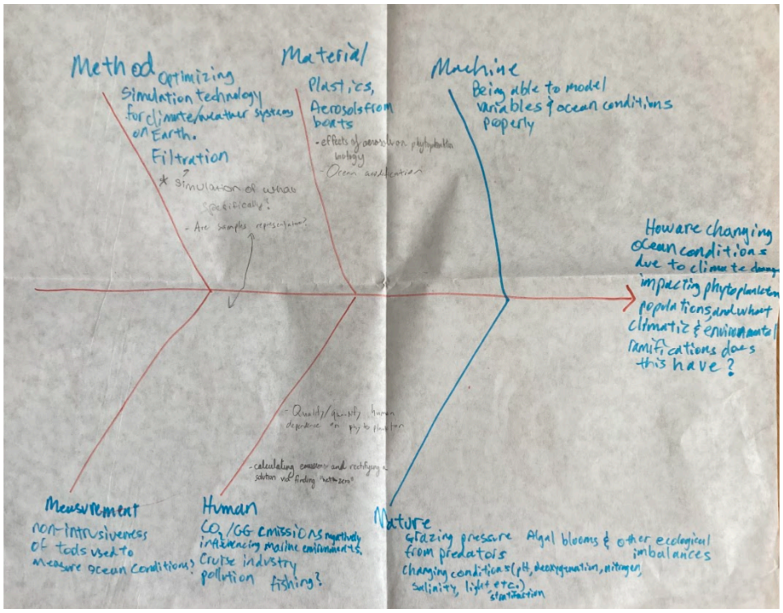
This is the mind map from my pie chart on climatology. Seeing as my biggest passion for my project is exploring phytoplankton, all ideas generated off of this pie chart relate to phytoplankton in one way or another. Fields ranging from biochemistry to gas and nutrient stoichiometry, meteorology, and systems modeling, and more are incorporated. Currently, the most intriguing areas that could be focused on include analyzing micronutrients, the impacts of pH, and factors influencing metabolism.

Fishbone Diagrams


METHODS
Monoculture reforestation vs. diverse
Frequency of adding compost to soil
↳ w/o damaging ecosystems?
SOIL
microbe health
presence of nitrogen-fixing bacteria
Can the number of nitrogen-fixing bacteria from compost be measured. How can that number be used?
Are there other threats to soil that can impact its health? Can bacteria or certain animals be used to help it?
Animals impact on soil and compost amount needed?
FOOD SUPPLY CHAIN
Effect of fertilizer & pesticide use in agriculture
Too much fertilizer / compost can cause harm to soil, how can you balance the usage?
↳ while preventing eutriphication?
How can ~~food waste~~ compost be harnessed to reduce food waste, and how can it be used to improve soil conditions for restoration efforts, given varying conditions across global ecosystems?
How can people living in cities be encouraged to compost when they don't have gardens?
↳ What do they do with compost? Used in local green areas? Encourage more personal gardens?
littering,
impact of pollution
↳ Can litter be used for compost?
Human Impact
Certain plants require dif. nutrients/ soil
↳ maybe have different composts/ nutrients that help soil
Can composting devices be made more efficient
↳ Not as smelly when opened would make it more appealing
Food used
C/N ratio
temperature
↳ Can different compost recipes be found for different environments?
Composting parameters
How can compost be properly distributed?
↳ Use animals? (like how plants spread seeds?)
- Must take into account dif. environments
Wetlands, peatlands
↳ water absorbing bacteria?
↳ bacteria that absorb certain nutrients, ones which whose roles can be utilized experimentally?
Reforestation/ Restoration
Abhinav K. Sharma

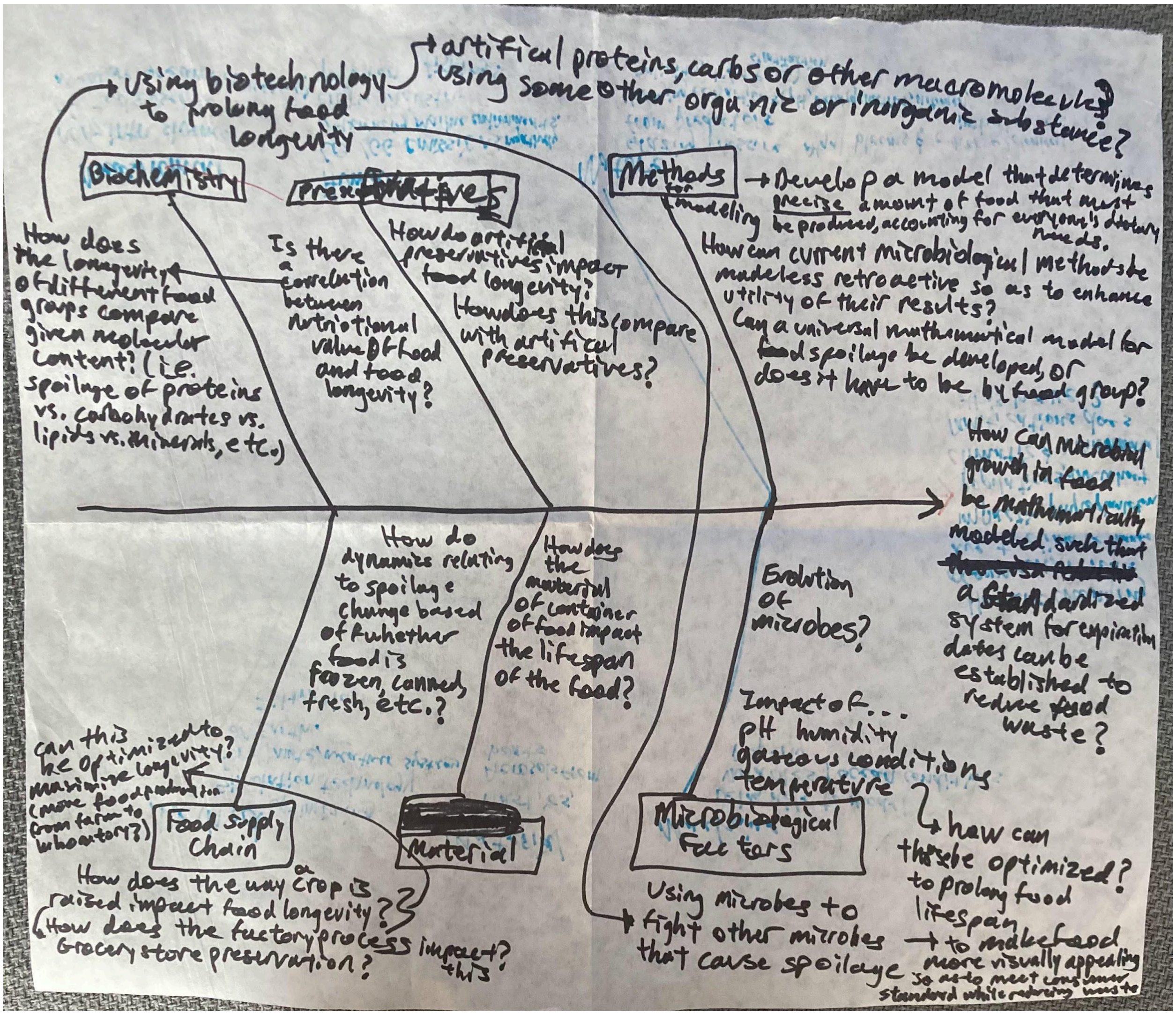
Abhinav K. Sharma 4 September 2023 2:56 PM

This is my fishbone-diagram for my idea relating to identifying ideal composts for different ecosystems across the globe while harnessing food waste in the process. The six sectors collapsing into this idea include reforestation/restoration (looking at the methods and specific needs for different ecosystems), the food supply chain (dealing with food waste specifically), composting parameters (how compost ought to be organized, its composition and such), soil (analyzing soil ecology and soil science for this topic), methods (evaluating efficacy of different reforestation models), and human impact (assessing anthropogenic impacts with regards to this issue).


Method
Optimizing Simulation technology for climate/weather systems on Earth.
Filtration
* Simulation of what specifically?
- Are samples representative?
Material
Plastics, Aerosols from boats
- effects of aerosols on phytoplankton biology
- Ocean acidification
Machine
Being able to model variables & ocean conditions properly
How are changing ocean conditions due to climate change impacting phytoplankton populations, and what climatic & environmental ramifications does this have?
- Quality/quantify human dependence on phytoplankton
- calculating emissions and rectifying a solution via finding "net zero"
Measurement
non-intrusiveness of tools used to measure ocean conditions?
Human
CO_2/GG emissions negatively influencing marine environments.
Cruise industry
pollution
fishing?
Nature
Grazing pressure from predators
Algal blooms & other ecological imbalances
changing conditions (pH, deoxygenation, nitrogen, salinity, light, etc.) stratification

Abhinav K. Sharma
4 September 2023 2:56 PM

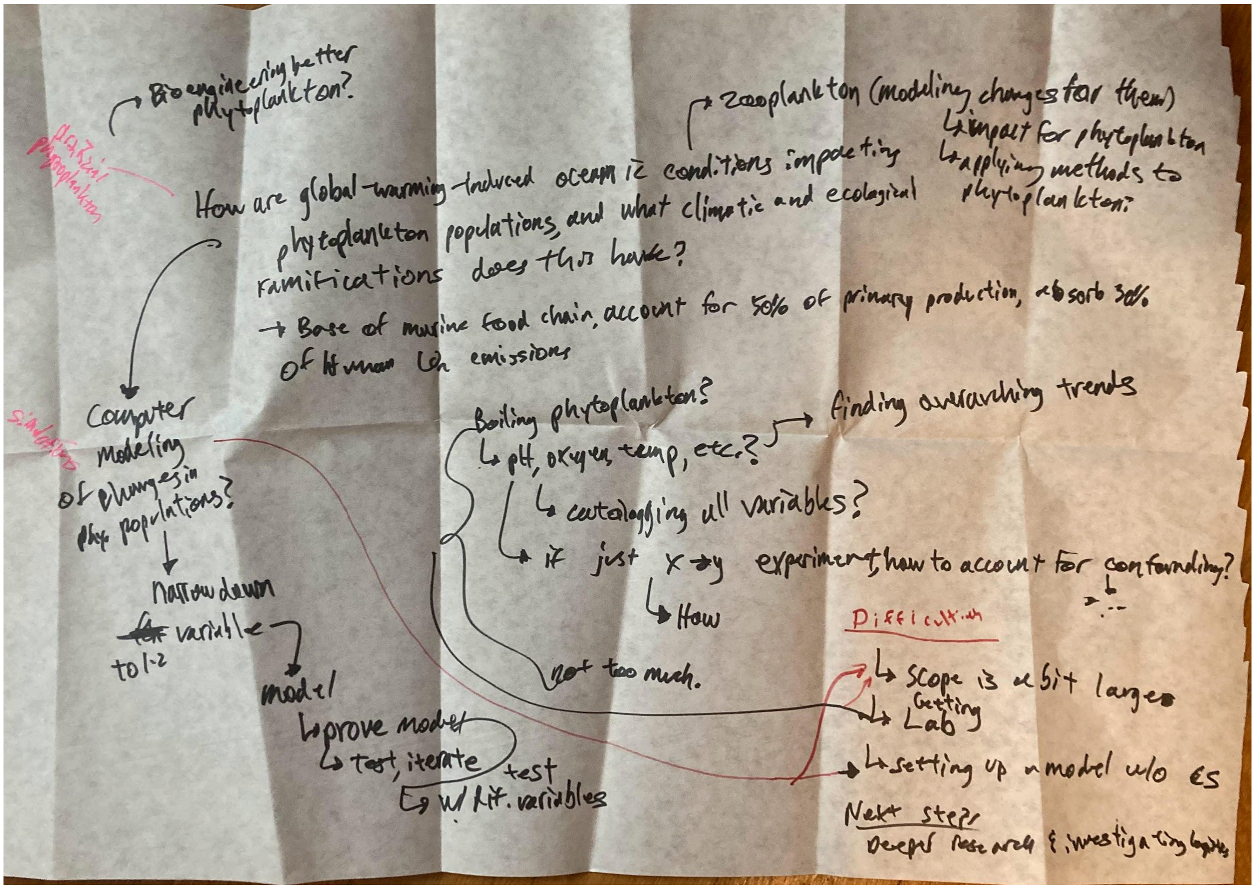
This fish-bone diagram is about the main project idea relating to phytoplankton. By focusing on the mechanical, material, methodological, anthropogenic, historical, mensurational aspects relating to phytoplankton, many various insights were reached. This includes recognition of the importance of modeling phytoplankton trends via computer software, as well as the various ecological and anthropogenic impacts. This process helped contextualize the study of phytoplankton throughout the entire process.


Using biotechnology to prolong food longevity
artificial proteins, carbs or other macromolecules using some other organic or inorganic substance?
Biochemistry
Preservatives
Methods for modeling
Develop a model that determines precise amount of food that must be produced, accounting for everyone's dietary needs.
How does the longevity of different food groups compare given molecular content? (i.e. spoilage of proteins vs. carbohydrates vs. lipids vs. minerals, etc.)
Is there a correlation between nutritional value of food and food longevity?
How do artificial preservatives impact food longevity?
How does this compare with artificial preservatives?
How can current microbiological methods be made less retroactive so as to enhance utility of their results?
Can a universal mathematical model for food spoilage be developed, or does it have to be by food group?
How can microbial growth in food be mathematically modeled such that ~~there is a~~ a standardized system for expiration dates can be established to reduce food waste?
How do dynamics relating to spoilage change based of whether food is frozen, canned, fresh, etc.?
How does the material of container of food impact the lifespan of the food?
Evolution of microbes?
Impact of... pH humidity gaseous conditions temperature
how can these be optimized? to prolong food lifespan
to make food more visually appealing so as to meet consumer standard while reducing waste
can this be optimized to maximize longevity? maximize food production (more food from farm to who eats it?)
Food Supply Chain
Material
Microbiological Factors
How does the way a crop is raised impact food longevity?
How does the factory process impact this?
Grocery store preservation?
Using microbes to fight other microbes that cause spoilage

Abhinav K. Sharma
4 September 2023 4:42 PM

This is the fish-bone diagram of my idea regarding modeling microbial growth in various food products as a means of developing a series of standardized formulas for creating a system for expiration dates, thereby reducing food waste from this sector. All six vectors that exude from these topics are fields of science and societal systems that are relevant to this issue. Methods consider the means and ways in which food is produced, transported, and stored. Similarly, the food supply chain analyzes these trends, while also considering factors such as food waste and CO_2 emissions, while the material sector considers the material used in the food supply process. Preservatives, biochemistry, and microbiological factors offer various lenses that strike at the direct science related to food and microbes.

7-Hats


Bioengineering better phytoplankton?
Artificial phytoplankton
How are global-warming-induced oceanic conditions impacting phytoplankton populations, and what climatic and ecological ramifications does this have?
Zooplankton (modeling changes for them)
impact for phytoplankton
applying methods to phytoplankton?
Base of marine food chain, account for 50% of primary production, absorb 30% of Human CO2 emissions
simulations
Computer modeling of changes in phy. populations?
narrow down ~~[illegible]~~ variable to 1-2
model
prove model
test, iterate
test
w/ dif. variables
Boiling phytoplankton?
pH, oxygen, temp, etc.?
finding overarching trends
catalogging all variables?
if just x → y experiment, how to account for confounding?
How
not too much.
Difficulties
Scope is a bit large
Getting Lab
setting up a model w/o Es
Next steps
Deeper Research & investigating logistics

Abhinav K. Sharma

10 September 2023 5:58 PM

7-Hats brainstorming method was employed during the Bournedale retreat. The topic used here was the ideas relating to modeling the impact of various oceanic conditions on phytoplankton (given global warming-induced change in oceanic conditions). In effect, this exercise was a meditation on the essence of experimental design. Namely, it was about not having confounding variables yet still properly simulating the ocean. This dilemma was resolved by developing two possible approaches to a project on phytoplankton: either computer simulation or lab experiments.

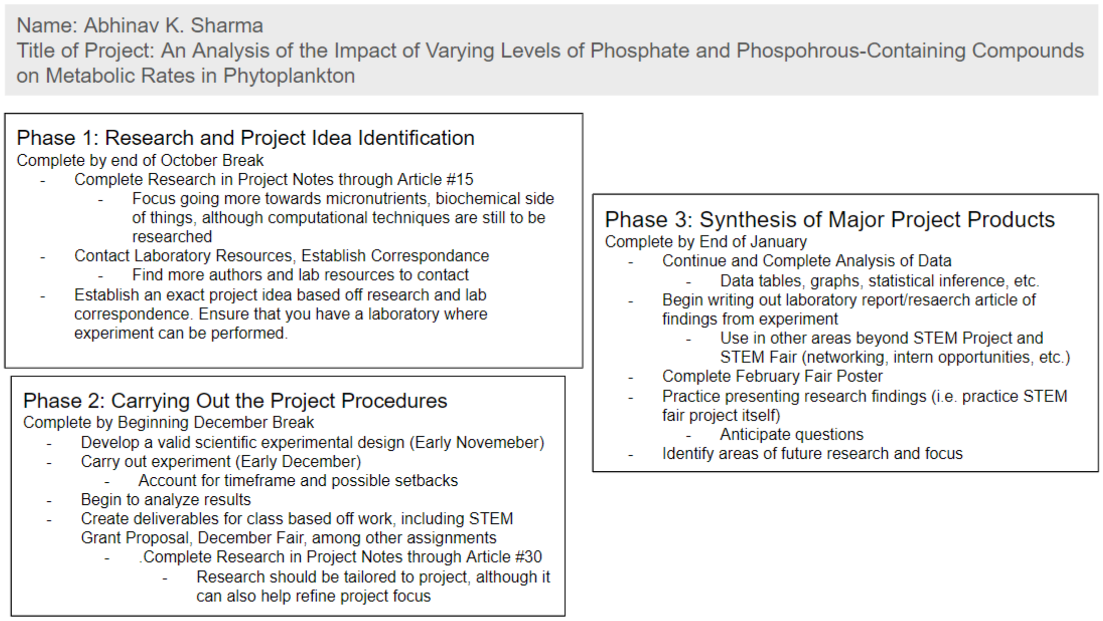
Project Introduction:

Research Question:

Hypothesis:

Brief Overview:
Project Outline (evolving research and project timeline is provided in daily entries below) Additionally, below are timelines created for the entire project:

Project Timeline: Version I



Name: Abhinav K. Sharma
Title of Project: An Analysis of the Impact of Varying Levels of Phosphate and Phospohrous-Containing Compounds on Metabolic Rates in Phytoplankton

Phase 1: Research and Project Idea Identification

Complete by end of October Break

- Complete Research in Project Notes through Article #15
    - Focus going more towards micronutrients, biochemical side of things, although computational techniques are still to be researched
- Contact Laboratory Resources, Establish Correspondance
    - Find more authors and lab resources to contact
- Establish an exact project idea based off research and lab correspondence. Ensure that you have a laboratory where experiment can be performed.

Phase 2: Carrying Out the Project Procedures

Complete by Beginning December Break

- Develop a valid scientific experimental design (Early Novemeber)
- Carry out experiment (Early December)
    - Account for timeframe and possible setbacks
- Begin to analyze results
- Create deliverables for class based off work, including STEM Grant Proposal, December Fair, among other assignments
    - .Complete Research in Project Notes through Article #30
        - Research should be tailored to project, although it can also help refine project focus

Phase 3: Synthesis of Major Project Products

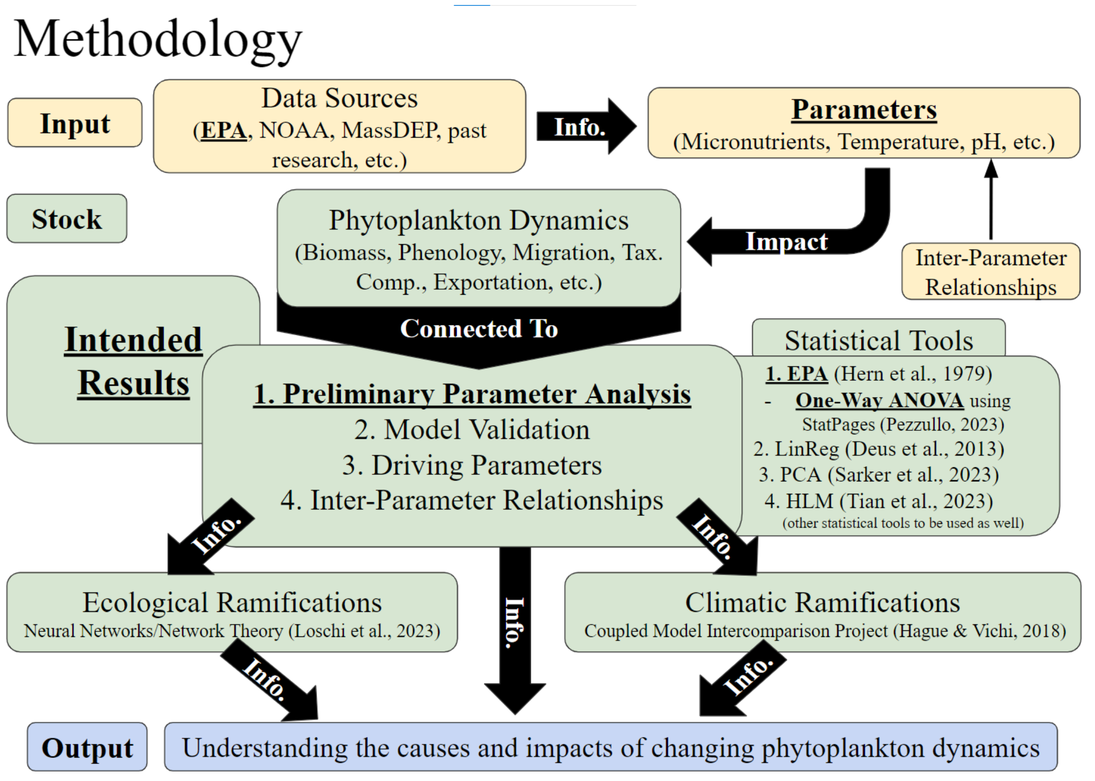
Complete by End of January

- Continue and Complete Analysis of Data
    - Data tables, graphs, statistical inference, etc.
- Begin writing out laboratory report/resaerch article of findings from experiment
    - Use in other areas beyond STEM Project and STEM Fair (networking, intern opportunities, etc.)
- Complete February Fair Poster
- Practice presenting research findings (i.e. practice STEM fair project itself)
    - Anticipate questions
- Identify areas of future research and focus

5:22 AM 10 October 2023 Abhinav K. Sharma

Above is my current STEM project plan as of 3 October 2023. It revolves around getting a laboratory space, developing a scientifically valid experimental design, performing that procedure, analyzing data, forming conclusions, and synthesizing all required products in a swift, efficient manner that meets all deadlines and keeps up with, if not, exceeds, the pace.

Project Overview (Timeline Version II)


Methodology
Input
Data Sources (EPA, NOAA, MassDEP, past research, etc.)
Info.
Parameters (Micronutrients, Temperature, pH, etc.)
Stock
Phytoplankton Dynamics (Biomass, Phenology, Migration, Tax. Comp., Exportation, etc.)
Impact
Inter-Parameter Relationships
Intended Results
Connected To
1. Preliminary Parameter Analysis
2. Model Validation
3. Driving Parameters
4. Inter-Parameter Relationships
Statistical Tools
1. EPA (Hern et al., 1979)
- One-Way ANOVA using StatPages (Pezzullo, 2023)
2. LinReg (Deus et al., 2013)
3. PCA (Sarker et al., 2023)
4. HLM (Tian et al., 2023)
(other statistical tools to be used as well)
Info.
Info.
Ecological Ramifications
Neural Networks/Network Theory (Loschi et al., 2023)
Info.
Climatic Ramifications
Coupled Model Intercomparison Project (Hague & Vichi, 2018)
Info.
Info.
Output
Understanding the causes and impacts of changing phytoplankton dynamics

7:14 AM 12 December 2023 Abhinav K. Sharma

The above is the systems diagram that has been developed to describe the methodology of this project. Bolded and underlined portions were the major parts completed for the December Fair. This framework provides a clear path to be followed for the project. The input of the system contains parameter data. Meanwhile, the stock consists of the planned computational methods. Finally, the output is the goal of this study.

Final Project Timeline and Computational Apparatus

September

- Brainstorming Project Ideas
- Establishment of Phytoplankton as an Area of Focus
- Establishment of Dual-Focus
    - Modeling Phytoplankton Biochemistry and Genomics Given Micronutrient Concentrations and other Environmental Factors

    - Impact on primary production, metabolic processes, among other areas
  - Computationally Modeling Phytoplankton Dynamics (biomass, phenology, migratory patterns, etc.)
    - Impact on climate and food web
    - Using environmental conditions as parameters.

October

- Narrowing Down Project Focus
  - Identifying Computationally Modelling as Primary Approach
- Researching on Computational Techniques, Software, Possible Sources of Data For Input
  - NetLogo
  - TensorFlow and other Artificial Intelligence Modelling and Machine and Deep Learning Tools
  - Network Theory, Neural Networks, Louvain Method,
  - EPA, MassDEP, and other potential databases

November

- Choosing Sources of Data; Data Investigated Include:
  - EPA
  - MassDEP
  - USDA
  - Past Research (specific articles with data can be found in daily entries below)
- Choose Potential Variables to Investigate
  - Oxygen
  - Temperature
  - pH
  - Salinity
  - Nutrients

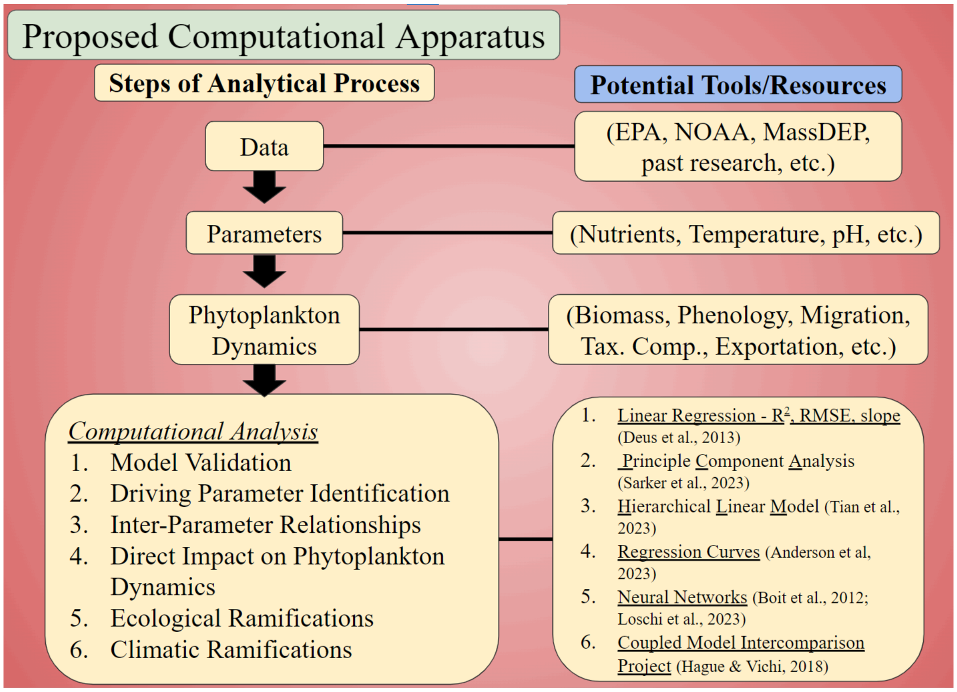
- Develop Initial Testing Strategy Using the Following Outline:
    - Preliminary Parameter Analysis
    - Model Validation
    - Driving Parameters
    - Inter-Parameter Relationships
    - Climatic Ramifications
    - Ecological Ramifications

December

- Create a Preliminary Statistical Analysis
    - ANOVA testing on homogeneity of parametric preferences among 57 Phytoplankton Genera (Experiment I)
- Continue Research on Potential Climate and Environmental Modeling tools
    - Research Articles
- Model Iteration

January

- Develop finalized models apparatus:



Proposed Computational Apparatus

Steps of Analytical Process

Potential Tools/Resources

Data — (EPA, NOAA, MassDEP, past research, etc.)

Parameters — (Nutrients, Temperature, pH, etc.)

Phytoplankton Dynamics — (Biomass, Phenology, Migration, Tax. Comp., Exportation, etc.)

Computational Analysis

1. Model Validation
2. Driving Parameter Identification
3. Inter-Parameter Relationships
4. Direct Impact on Phytoplankton Dynamics
5. Ecological Ramifications
6. Climatic Ramifications

1. Linear Regression - R^2, RMSE, slope (Deus et al., 2013)
2. Principle Component Analysis (Sarker et al., 2023)
3. Hierarchical Linear Model (Tian et al., 2023)
4. Regression Curves (Anderson et al, 2023)
5. Neural Networks (Boit et al., 2012; Loschi et al., 2023)
6. Coupled Model Intercomparison Project (Hague & Vichi, 2018)

9:46 AM 11 February 2024 Abhinav K. Sharma

This is the final version of the proposed computational apparatus in studying the causes and effects of the changing nature of phytoplankton dynamics. First, empirical data is attained, providing information on parametric values. The value of environmental parameters exerts some impact on phytoplankton populations. This process has underlying causes and effects which can be characterized through following the six step modeling process described above.

- Identify future areas of focus

February

- Develop necessary products (Research Paper, Documentation, Poster, etc.)
- Present results

Professional Communication:

Emails made prior to project development (September and October):

(This individual was emailed as their article provided important information about modeling zooplankton. The hope was that in reaching out to them, insights about modeling phytoplankton could be derived)

To: L.Ratnarajah@liverpool.ac.uk

Lead-Author of "Monitoring and modelling marine zooplankton in a changing climate" (Article #7)

https://www.nature.com/articles/s41467-023-36241-5.pdf

Subject Line: Questions Regarding Recent Publication (Monitoring and modelling marine zooplankton in a changing climate)
Greetings, Dr. Ratnarajah.

I am Abhinav K. Sharma, a student at the Massachusetts Academy of Math and Science (MAMS) at the Worcester Polytechnic Institute (WPI). I am currently working on developing a five-to-six-month research project on phytoplankton for a science fair. Specifically, my project relates to modeling the impact of changing ocean conditions on phytoplankton at a global scale, taking into account regional variations, and various biotic and abiotic conditions. In conducting background research, I have read your article entitled "Monitoring and modeling marine zooplankton in a changing climate" from Nature Communications published earlier this year.

I have found that the contents of the article were incredibly helpful in developing my project. Content wise, it was easy to understand both the overall trends zooplankton are undergoing, and the necessary future steps needed for sampling and researching them. The insights offered have helped me make connections to my project regarding phytoplankton.
That being said, I wanted to ask a few questions regarding your article:

One of the current limitations with research in the field that was cited in your article is lack of resources to model zooplankton dynamics and the factors that cause that. Do similar problems exist when it comes to modeling phytoplankton?

How are phytoplankton impacted by the trends in zooplankton populations outlined in your article? Indeed, many of the findings in your article about zooplankton have similarities to what I have found in my research regarding phytoplankton, for example, the preference in smaller cell size.

As such, in researching zooplankton populations, I hope to make important connections to phytoplankton as I work on my project. If you could offer any additional insights, that would be much appreciated. Additionally, if you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

The author for Article #8 was contacted due to the numerous interesting biochemical connections to and ramifications for phytoplankton it provided. More insight and information was desired. Hitherto he has been the only correspondence that has responded to me.
To: pengjin@gzhu.edu.cn

Lead-Author of “DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change” (Article #8)

Subject Line: Questions and Interest Regarding Recent Publication (DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change)

Email Body:

Greetings, Dr. Jin.

I am Abhinav K. Sharma, a student at the Massachusetts Academy of Math and Science (MAMS) at the Worcester Polytechnic Institute (WPI). I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project relates to investigating the impact of how changing ocean conditions are impacting phytoplankton. In conducting research, I have read your article entitled “DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change” published earlier this year in the Journal of Experimental Botany.

I have found that the contents of the article were incredibly helpful in developing my project. It has provided important insights about the genetic and biochemical dynamics that phytoplankton populations are facing. I find the possibility of DNA methylation being a possible means of adaptation to be truly fascinating. These ideas have been helpful in developing my project regarding phytoplankton.

That being said, I wanted to ask a few questions regarding your article:

Beyond temperature and CO_2 concentration, are there other variables (for example, dissolved oxygen or salinity), that impact gene regulation, and thereby metabolism in phytoplankton?

Are there other biochemical adaptations beyond DNA methylation that phytoplankton can use? Do these methods have any connection to DNA methylation?

How do diversions from the Redfield Ratio impact the rate of DNA methylation in phytoplankton? How does this impact metabolic rates?

If you could offer any additional insights, that would be much appreciated. Additionally, if you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

Reply:

Dear Abhinav K Sharma,

Thank you very much for your message and your interests on our paper. Please find my responses below your questions. I hope my responses may be helpful for you.

Best regards,

Peng

Beyond temperature and CO2 concentration, are there other variables (for example, dissolved oxygen or salinity), that impact gene regulation, and thereby metabolism in phytoplankton?

Yes, for sure. There were a large body of literature looking at the impacts of environmental factors (such as nutrient, DO and salinity as you indicated) on the gene transcriptions of phytoplankton.

Are there other biochemical adaptations beyond DNA methylation that phytoplankton can use? Do these methods have any connection to DNA methylation?

My answer to this question is also yes. Phytoplankton may alter their biochemical compositions of fatty acids, amino acids, phenolic compounds, etc, in responding to environmental changes. DNA methylation regulates the gene transcription (as we demonstrated in the JEB paper), and thereby regulate the metabolism of biochemical compounds. So there are expected to be tightly coupled.

How do diversions from the Redfield Ratio impact the rate of DNA methylation in phytoplankton? How does this impact metabolic rates?

You raised a good point here. But to the best of my knowledge, there are no study investigating the relationship between Redfield Ratio and DNA methylation.

Vidali was emailed as he works with plants, i.e. autotrophic organisms. With his work focusing on autotrophic organisms, the hope was that he could offer some vital insights into the connections between phytoplankton and other plant species.

Luis Vidali (studies plants, member of Biology and Biotechnology Department at WPI):
Subject Line: Questions Regarding Research from a Mass Academy Student

Email Body:
Greetings, Dr. Vidali.

I am Abhinav K. Sharma, a student from Mass Academy at WPI. I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project aims to investigate the impact of global warming-induced oceanic changes on phytoplankton at a biochemical level. As part of my project development, I have investigated some of your work due to its focus on autotrophic organisms.

I notice that your research centers around the cellular dynamics of plants with a particular focus on the cytoskeleton. With significant experience with the cytology and biochemistry of autotrophic organisms, I was wondering if you could answer some of the following questions:

How applicable are the findings of the model organism *Physcomitrella patens* to marine autotrophs, including phytoplankton? Is it too taxonomically separate from them for findings to be extrapolated?

Among the autotrophs that you have worked with, how do different rates of cellular growth, communication, and transportation brought about by varying conditions in the cytoskeleton impact metabolic and photosynthetic rates?

If you could offer any additional insights, that would be much appreciated. Additionally, I am interested in learning more about your research in depth. If you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

This individual was contacted due to their work with biochemistry and specific focus on phosphorus-containing biomolecules. Hope is to gain more insight on the impact of this specific micronutrient, extrapolating that to other similar relationships, and even getting a lab opportunity.
Arne Gerike (works with proteins and lipids):
Subject Line: Questions Regarding Research from a Mass Academy Student

Email Body:
Greetings, Dr. Gerike.

I am Abhinav K. Sharma, a student from Mass Academy. I am currently working on developing a five-to-six-month research project regarding phytoplankton for a STEM science fair. Specifically, my project aims to investigate the impact of global warming-induced oceanic changes on phytoplankton at a biochemical level. In particular, I have considered investigating how varying concentrations of phosphates and other phosphorus-containing compounds impact metabolic rates in phytoplankton. As part of my project development, I have investigated some of your work due to its focus on interactions between proteins and lipids.

In particular, I notice that you have a particular focus on phosphoinositides and their various interactions with proteins. With significant experience with phosphorus-containing molecules and biochemical interactions, I was wondering if you could answer some of the following questions:

What factors control the type and amount of phosphoinositide lipids found in a cell?

How does phosphoinositide composition and concentration differ across different organisms?

Similarly, how do protein interactions with phosphoinositide lipids differ among terrestrial and aquatic organisms? Among autotrophic and heterotrophic organisms?

If you could offer any other insights, that would be much appreciated. Additionally, I am interested in learning more about your research in depth. If you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

This individual was emailed because their work at UConn relates very well to my project focus on phytoplankton. They have a lot of experience with the ecology and biochemistry of this species, and they focused on phosphorus-containing compound phytic acid, which connected with my specific interest. So, due to these good connections, and the possibility of remote lab work, I reached out to this individual.
Subject Line: Questions and Interest Regarding Recent Publication (Two-sided effects of the organic phosphorus phytate on a globally important marine coccolithophorid phytoplankton)
Email to Dr. Lin:

I am Abhinav K. Sharma, a student at the Massachusetts Academy of Math and Science (MAMS) at the Worcester Polytechnic Institute (WPI). I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project relates to investigating the impact of how changing ocean conditions are impacting phytoplankton at a biochemical level. In particular, I have considered investigating how varying concentrations of phosphates and other phosphorus-containing compounds impact metabolic rates in phytoplankton. In conducting research, I have read your article entitled “Two-sided effects of the organic phosphorus phytate on a globally important marine coccolithophorid phytoplankton” published just last month in the Microbiology Spectrum Journal.

I have found that the contents of the article were incredibly helpful in developing my project. It has provided important insights regarding the role phytic acid plays in upregulating many metabolic processes, but also its possible toxic effects. These ideas have been helpful in developing my project regarding phytoplankton. I also notice that you are from UConn, and work a lot with the ecology and biochemistry of phytoplankton and marine environments.

That being said, I wanted to ask a few questions regarding your article and work:

How do the varying cytological attributes of different phytoplankton species influence the optimum phosphate levels under which they can operate?

Does an enhancement in metabolic processes among phytoplankton result in an enhancement in metabolic processes among other organisms higher up the trophic pyramid? How do varying metabolic rates impact symbiotic activity in phytoplankton?

Have any of your PhD students worked remotely with laboratory-collected data, specifically DNA or molecular data?

If you could offer any additional insights, that would be much appreciated. Additionally, if you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

This individual was emailed because of their connections with computer modeling of complex ecological systems (bee pollination dynamics). She was contacted with the hope of gaining insights about how the NetLogo software works. With the direction of my project heading towards a computational model of global migratory patterns of phytoplankton given global warming-induced oceanic changes, and the resulting ecological and climatic impacts, reaching out to this individual as imperative.

Email to Dr. Ryder:
Greetings, Dr. Ryder.

I am Abhinav K. Sharma, a student from Mass Academy here at WPI. I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project aims to investigate the impact of global warming-induced oceanic changes on phytoplankton. In doing so, I plan to utilize computer modeling. As part of my project development, I have investigated some of your work due to its implementation of computational techniques.

I notice that your work has involved using NetLogo to create variables that were then used to model bees and their pollination patterns, and that your class, Simulation in Biology, involves students building their own simulations of biological systems. Given this experience, I was wondering if you could answer some of the following questions:

For someone who has only a very basic background in computer science, how steep of a learning curve does NetLogo pose?

Do you believe it is possible to create predictive ecological and climatic models of oceanic systems at a global scale, taking into account the impact of global warming-induced ocean changes on phytoplankton? What level of complexity can programs on NetLogo be?

Have any of your past students worked in any area that relates to what I mentioned in the previous question?

If you could offer any additional insights, that would be much appreciated. Additionally, I am interested in learning more in-depth about your work. If you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

Emails made after project development (February):

This organization was contacted, given past correspondence and local ties. It was hoped that the tools created as a result of this project could be benevolently applied. To this end, I reached out to an organization in my hometown to try and get this process started.

Greetings, Westford Climate Action Group!

I am Abhinav K. Sharma, a high school junior currently attending the Massachusetts Academy of Math and Science (MAMS), sponsored by WPI. To clarify, I am still a Westford resident and previously attended Westford Academy. I have been on the emailing list for this organization since December 2022. In reading the newsletters, it has been excellent to see the efforts this organization engages in. The dedication to a healthier, more sustainable Westford is clear.

I have also seen this organization at various public events, including the 2022 WA Holiday Bazaar (when I originally signed up to receive updates), the Earth Day Celebration 2023, as well as the 2023 WA Holiday Bazaar. While speaking with the representative at the most recent Holiday Bazaar, I was told that this organization was interested in recruiting students and younger members, and that if interested in joining, I should reach out. When I mentioned that I was from MAMS, I was also told that my membership might prove useful, given the potential resources and opportunities that could be provided for this organization.

At this time, I am interested in potentially becoming a member of the Westford Climate Action group. I am passionate about fully addressing the climatic and environmental crises our planet faces, and plan on pursuing a degree in environmental and/or climate science in college. However, given a tight schedule, I would first like to get to know what responsibilities my membership would entail.

Moreover, I would like to present a major project that I have been working on as part of the MAMS program. In this research project, I developed a system of computer models capable of delineating the causes and effects of changing phytoplankton populations. Phytoplankton play a major role in climate regulation and are indicators of ecosystem health. At a local level, understanding the impact of their populations on lakes, streams, and other bodies of water is crucial for developing a better understanding of overall ecosystem health. The model that I created has the potential to be a useful tool for this end.

I am choosing to present this tool to Westford Climate Action in the hopes that it can be used to help optimize aquatic ecosystem health, in turn helping the organization in its overarching goal of a healthier local environment. I am considering presenting this to the Westford Water Department as well. I have attached a draft of my research paper describing this project in greater detail. However, I am also currently developing other mediums for presenting this project, including various posters and slideshows as part of other school activities.

In addition to joining this organization, I would greatly appreciate it if I could be provided the opportunity to present my work at a future meeting. I am also interested in participating with this organization's plans for this year's Earth Day Celebration, and potentially presenting my work there as well, if possible.

I hope that my offers are considered and that I can become a member of this organization, helping in their goals for a greener Westford. Thank you for the consideration, and apologies for the lengthy email.

Thank You Very Much,
Abhinav K. Sharma

My mentor had made me aware that there have been conversations at the regional (New England) level about how to address proper lake/pond/overall water-body management. Therefore, after finishing the creation of the first major iteration of this computational apparatus as part of my project, I decided that emailing the MassDEP staff would be worth a try, to see if my tool could be applied in a viable manner.

Greetings, DRD Divris!

I am Abhinav K. Sharma, a high school junior currently attending the Massachusetts Academy of Math and Science (MAMS), sponsored by WPI. However, as a Westford resident, I am a part of the Northeast region of the MassDEP. Thus, I am choosing to reach out to you regarding a potential modeling tool for aquatic ecosystem health.

This research project presents a unified system of multiple computer and statistical models that delineate the environmental and climatic causes and effects of changing phytoplankton populations. At a high level, a wide variety of modeling tools are arranged in a methodical order to achieve this end. Given the importance phytoplankton play in ecosystem health and climate regulation, understanding the nature of their populations is crucial for better understanding environmental and climatic systems at a broader level.

This project was carried out as part of the MAMS program. While I have attached a draft of my research paper for this project, I am currently working on developing other mediums for presenting this project as part of other school activities and can provide those documents upon request.

I am presenting this paper in the hopes that this may be a potential tool in assisting the MassDEP in managing marine ecosystems and securing high water quality. I would greatly appreciate it if an opportunity for further correspondence on the potential use of this model could be offered. Alternatively, if there is another individual better suited for discussing this topic, I would gladly appreciate it if I could be directed towards them.

Thank You Very Much,

Abhinav K. Sharma

Materials and Methods:

Role of Student vs. Mentor

The author of this paper was the student, who was mentored by Dr. Kevin Crowthers. From July 2023 to February 2024, the bulk of the work, including idea generation and attainment, research, and model development, validation, and testing, was performed by the former. The latter was responsible for monitoring project progress, as well as offering advice, particularly on potential software usage for parametric testing.

Equipment and Materials

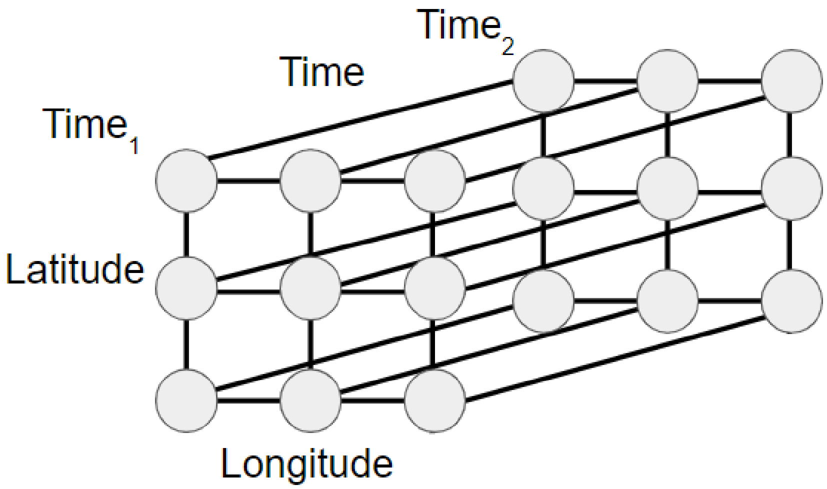
The primary dataset analyzed in this study was the 2018 World Ocean Database (WOD18) provided by the National Oceanic and Atmospheric Administration (NOAA). Both spatially and temporally, this dataset provides a highly cosmopolitan measurement of numerous environmental parameters, including water temperature, micronutrients, pH, salinity, among many others (Boyer et al., 2018). Access to the files of this dataset was attained through the Registry of Open Data provided by Amazon Web Services (AWS). The files used in this dataset were all updated within the AWS S3 explorer system on 17 October 2023 when obtained. Files were organized by year from 1900 to 2023, with pre-1900 data being referred to as 1800 (Amazon Web Services, 2024). The file of each year was systematically downloaded. Regression models were developed using Python code. The main programming interface used was Google Colaboratory, which, when used, was most recently updated on 8 January 2024, supporting Python 3.10.12 (Google Colaboratory, 2024). An online webpage was used to conduct PCA (Statistics Kingdom, 2017). Across all Python programs developed, the Pandas, Xarray, NumPy, SciPy, SciKitLearn, Seaborn, and MatPlotLib packages were utilized. Additionally, to observe the dimensions of the data files more closely, the Panoply software, provided by the National Aeronautic and Space Administration’s (NASA’s) Goddard Institute for Space Studies (GISS), was downloaded. The most current version, 5.3.1, was used, released 1 January 2024 (NASA GISS, 2024).

Decisions for Parametric Model Development Based off Data Structure

Two key observations were made during preliminary use and analysis of the dataset that led to two key decisions on the analytical procedure performed. The first observation relates to the data structure of the files used. These were NetCDF (.nc) files, which illustrate parametric measurements at specific points of latitude, longitude, and depth across a series of equal temporal increments (Figure 1). In the metadata attained for the .nc files, these spatiotemporal attributes were referred to as “coordinates.” However, it was specifically noted within this directory that all dimensions, with the exception of the numerical measurements of parametric data, had their latitude, longitude, depth, and time alongside their measurement. Meaning, outside of the year as provided by the file, the spatiotemporal attributes of the primary data on environmental parameters could not be accessed. Therefore, a more holistic approach to analysis was taken, wherein the average global value of each factor was calculated for each year of data. This helped preserve temporal analysis and offered a viable overview of environmental relationships, though at the expense of omitting the nuances and consequences caused by dissimilar spatial attributes among parametric data.

Figure 1

Illustration of the NetCDF Data Structure

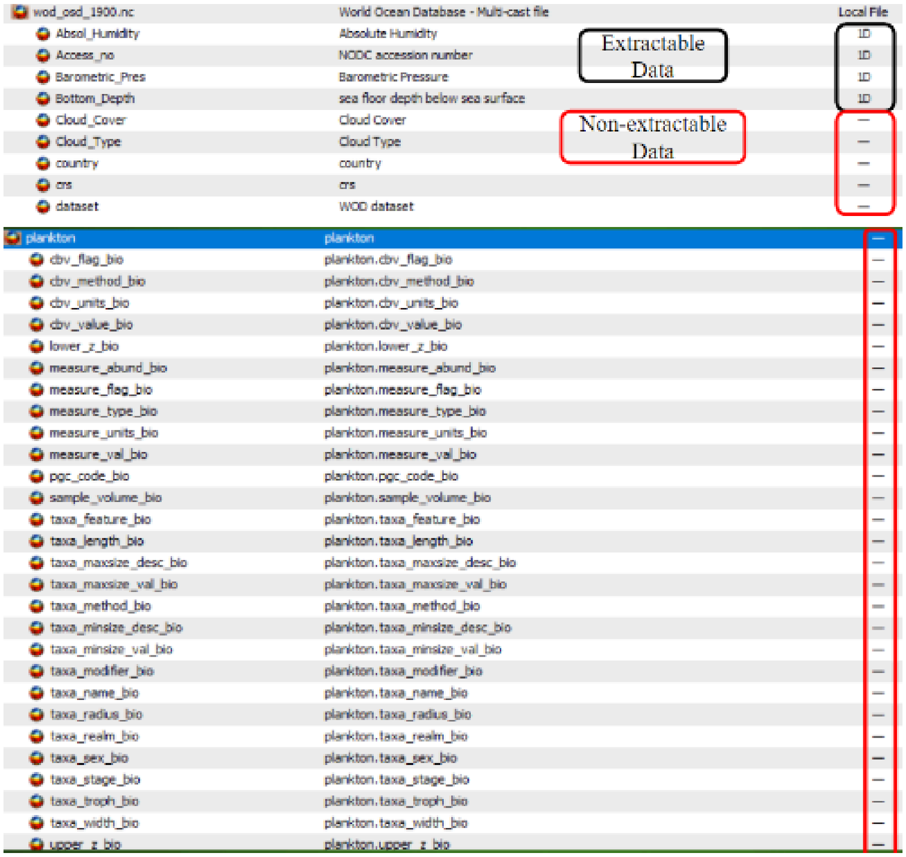


Note. This is a common illustration of the structure of a NetCDF file. The gray circles represent specific points in longitudinal and latitudinal space. Though not illustrated above, each of these coordinates also contains varying levels of depth. At all of these points in space, there exist parametric measurements. These measurements are projected out through a series of time increments, forming a three-dimensional figure.

The second observation made from the metadata was that planktonic data was inaccessible. Within the Panoply software, whereas extractable data was stored within one-dimensional arrays, the values of non-extractable data were unavailable. Figure 2 contains an illustration of this issue. Planktonic data fell under the latter category. Consequently, it was decided to use average global oceanic concentration of total chlorophyll as an indicator of phytoplankton dynamics, namely primary production. This is because chlorophyll is a crucial pigment for carrying out the photosynthetic process, and in turn, all other metabolic processes. As such, a higher concentration of chlorophyll would indicate greater potential for primary production, whereas Though a valuable indicator, it is important to note that it is not a direct measurement of phytoplankton traits. Although data of all available parameters spanning 1900 to 2019 were downloaded, due to the limited temporal range of measurements for total oceanic chlorophyll, this study focused on data from 1954 to 2017. This also reduced the number of factors assessed.

Figure 2

Inaccessibility of Planktonic Data



wod_osd_1900.nc	World Ocean Database - Multi-cast file	Local File
Absol_Humidity	Absolute Humidity	1D
Access_no	NODC accession number	1D
Barometric_Pres	Barometric Pressure	1D
Bottom_Depth	sea floor depth below sea surface	1D
Cloud_Cover	Cloud Cover	—
Cloud_Type	Cloud Type	—
country	country	—
crs	crs	—
dataset	WOD dataset	—
plankton	plankton	—
cbv_flag_bio	plankton.cbv_flag_bio	—
cbv_method_bio	plankton.cbv_method_bio	—
cbv_units_bio	plankton.cbv_units_bio	—
cbv_value_bio	plankton.cbv_value_bio	—
lower_z_bio	plankton.lower_z_bio	—
measure_abund_bio	plankton.measure_abund_bio	—
measure_flag_bio	plankton.measure_flag_bio	—
measure_type_bio	plankton.measure_type_bio	—
measure_units_bio	plankton.measure_units_bio	—
measure_val_bio	plankton.measure_val_bio	—
pgc_code_bio	plankton.pgc_code_bio	—
sample_volume_bio	plankton.sample_volume_bio	—
taxa_feature_bio	plankton.taxa_feature_bio	—
taxa_length_bio	plankton.taxa_length_bio	—
taxa_maxsize_desc_bio	plankton.taxa_maxsize_desc_bio	—
taxa_maxsize_val_bio	plankton.taxa_maxsize_val_bio	—
taxa_method_bio	plankton.taxa_method_bio	—
taxa_minsize_desc_bio	plankton.taxa_minsize_desc_bio	—
taxa_minsize_val_bio	plankton.taxa_minsize_val_bio	—
taxa_modifier_bio	plankton.taxa_modifier_bio	—
taxa_name_bio	plankton.taxa_name_bio	—
taxa_radius_bio	plankton.taxa_radius_bio	—
taxa_realm_bio	plankton.taxa_realm_bio	—
taxa_sex_bio	plankton.taxa_sex_bio	—
taxa_stage_bio	plankton.taxa_stage_bio	—
taxa_troph_bio	plankton.taxa_troph_bio	—
taxa_width_bio	plankton.taxa_width_bio	—
upper_z_bio	plankton.upper_z_bio	—

Note. The extractable data (black) were stored as one-dimensional arrays, whereas the values of inextricable data could not be obtained. All planktonic data fell under the latter category.

Data Extraction and Cleaning

Using Google Colaboratory, a brief program was written to extract parametric data .nc files and save them as Comma Separated Value (.csv) files. Within each .csv file, the average, standard error, and sample size for each year of each parameter was calculated and compiled into a separate spreadsheet file. The main data cleaning involved the removal of non-numerical data within .nc files when converting them to .csv. This was achieved by the Python program written.

However, for total oceanic chlorophyll and alkalinity, the data processing was more complex. When originally creating a time series for the former, it was noted that model strength was inhibited by abnormally high measurements around the early 2000s. Upon further investigation of the .csv files, it was noted that this was due to the abnormally high amount of outliers. In order to maximize model fitness, for chlorophyll data spanning 1998 to 2008, any and all measurements in excess of 20 µg/L were removed, and new averages, standard deviations, and sample sizes were determined. The next iteration of the time series had a stronger fit as a result. For alkalinity, many years had errors in how the data was recorded, in that the decimal place was improperly positioned. This led to values that were orders of magnitude too high for the dataset, and in turn, skewed summary statistics. As such, any such data was eliminated from the set, with summary statistics adjusted accordingly. Besides the processes described, all parametric data from 1954-2017 was preserved when performing data analysis.

Statistical Analysis

A variety of statistical tests and computational tools were used for the three major sets of models developed for this study. The time series for each parameter developed was created primarily using sinusoidal regression. The strength of each regression model was measured using a Pearson's Correlation, including both r and R^2. To assess correlation between each factor and total oceanic chlorophyll, linear regression, in conjunction with a Student's t-test for relationship significance and Pearson's correlation for relationship strength, was used. Finally, driving parameters were identified using PCA, along with supplementary techniques.

Sinusoidal Regression Including Pearson Correlation

Environmental features tend to be periodic in nature. The sine and cosine functions provide an effective way to model cyclical trends. Therefore, this specific type of a regression model was chosen, with R^2 and r measuring model strength and accuracy. This allowed for the evaluation of the validity of the overall computational system as well as projection abilities. Equation 1 represents the template function used for all time series models:

$$f(\kappa) = Asin((2\pi\gamma)\beta + \varepsilon) + \phi \tag{1}$$

Where $f(\kappa)$ is the function for the total chlorophyll concentration κ, A is the amplitude, 2πγ represents the length of the period (using radians, γ alone being in degrees), β is the given environmental factor (the next section enumerates the variable designation of each parameter), ε is the phase shift, and ϕ is the offset.

Additionally, using the offset as a midline for the sinusoid and the amplitude as a sort of ruler, an interval of all measurements projected by the sinusoid of each parameter was developed. Equation 3 represents the basic construction of the described sinusoidal interval:

$$\phi \ \pm A \tag{2}$$

Linear Regression Including Pearson Correlation and Student's t-test

For the relationship of every parameter with the indicator, total chlorophyll concentration, a Linear Regression model was developed. On a functional level, assessing each individual parameter's relationship with total chlorophyll acted as a precursor to identifying which of them had a significant influence on chlorophyll when all parameters were considered in tandem. In essence, performing linear regression acted as a prerequisite for then performing PCA. The significance of relationships were determined using a Student's t-test for Linear Regression at $\alpha = 0.05$. Both double- and single- tailed p-values were attained. This was done in conjunction with the use of R^2 and r to measure model accuracy and strength. Similar to the previous set of models, Equation 3 provides a template wherein each parameter's regression model was represented:

$$f(\kappa) \ = \ m\beta \ + \ \beta_0 \tag{3}$$

Where $f(\kappa)$ is the function for the total chlorophyll concentration κ, m is the predicted slope of the line, β is the given parameter, and β_0 is the y-intercept of the model.

PCA

In order to identify the driving parameters behind total chlorophyll concentrations, PCA was used. PCA is a dimension reduction technique that compresses multiple independent variables into fewer dimensions so as to summarize overarching data patterns and allow for ease of data visualization. Before performing PCA, the data must be standardized so that scale does not impede the accuracy of results. In this study, before PCA was performed, all data of the indicator (chlorophyll) and every parameter were standardized using minimum-maximum normalization. Within the setting of PCA, the total variance of the data is measured. This variance is captured by a finite set of portions of the data known as principal components. In two-dimensional representations, the first two principal components, that is, the two components that account for the highest amount of variance, denoted PC_1 and PC_2, are placed on the horizontal and vertical axes respectively. Since information from the other principal components (PC_3, PC_4, … PC_n) is omitted, it is important that most variance is captured by PC_1 and PC_2. This is measured by each principal component's eigenvector values, which are derived from various matrix operations performed on the data.

Then, to standardize the amount of variation each principal component captures, the eigenvalues are divided by the total variance of the dataset. A scree plot is used to depict the cumulative coverage of variance by all principal components. Along with a PCA plot, a scree plot was used to illustrate these notable properties of the principal components. In two-dimensional PCA, every parameter assessed captures some amount of either principal component, and holds either a positive or negative relationship with the directionality of component variances. This magnitude and directionality is represented by a pair of coordinates that form a vector. The greater the magnitude of variance represented by a parametric vector, the more influential that parameter is relative to the overall data, and in turn, driving the dependent variable. For the study, the magnitude of each parameter was calculated as depicted by Equation 4:

$$M(\beta) = \sqrt{(C_{PC1})^2 \cdot \nu_{PC1} + (C_{PC2})^2 \cdot \nu_{PC2}} \quad (4)$$

Where $M(\beta)$ is the function of the magnitude of parameter β, C_{PC1} is the contribution of the parameter to the variance of PC_1, while C_{PC2} is the contribution of the parameter to the variance of PC_2, and ν_{PC1} is the proportion of the total variance represented by PC_1, and ν_{PC2} is the proportion of the total variance represented by PC_2.

The magnitudes of each parameter were calculated using the above equation, and then subsequently ranked by descending magnitude values. The parameters with the highest calculated magnitude were identified as driving parameters of chlorophyll concentrations. Additionally, to assess the presence of inter-parameter relationships, a covariance matrix was used. A covariance matrix is an intermediate operation performed in the complex matrix calculations involved with PCA wherein all independent variables are arranged in a square array. The cells of this matrix contain the covariance between the row and column parameters. Values vary between 0 and 1, and can be either positive or negative based on the directionality of the relationship. A greater magnitude indicates a stronger relationship between the two parameters. The diagonal cells represent the variance of that individual parameter following dimension reduction procedures. The results from these three sets of computational models were then put into biogeochemical, ecological, and climatic context.

Background:

Computational Modeling of Phytoplankton Dynamics with Climatic and Ecological Ramifications

Phytoplankton encompass a broad range of aquatic, microscopic, photosynthetic species of viruses, bacteria, fungi, protists, animals, and archaea. They are responsible for about half of all global primary production, the production of nutritional organic matter from inorganic compounds via photosynthesis and other metabolic processes (Käse & Geuer, 2018). Phytoplankton are key to biogeochemical cycling, helping circulate nitrogen, phosphorus, silica, and other micronutrients (Sarker et al., 2023). They also absorb 30% of anthropogenic carbon emissions (Rohr et al., 2023). Beyond photosynthesis, carbon sequestration is also performed through exportation, a process where, after death, cellular matter sinks to the ocean floor, forming carbon sinks. Phytoplankton regulate climate not only through controlling carbon circulation, but also through light reflection. Certain functional groups produce dimethylsulfoniopropiothetin, a complex, sulfur-containing molecule. This compound decomposes into dimethylsulfide, which in turn decomposes into compounds that reflect solar radiation (Deppeler & Davidson, 2017). It is in fact believed that biochemical processes such as this one helped cause the first major ice ages on Earth (Käse & Geuer, 2018). Additionally, phytoplankton lie at the base of marine food chains, serving as prey for various species of zooplankton and fish (Käse & Geuer, 2018; Loschi et al., 2023). Therefore, phytoplankton are an integral part of the global climate and environmental systems, making the ability to understand how their operations and functionalities are to change because of global warming incredibly crucial.

Understanding The Impact of Global-Warming Induced Aquatic Changes on Phytoplankton

With that in mind, the impact global warming has had on oceanic conditions themselves must first be considered. Climate change has led oceans to becoming warmer, more acidic, anoxic, and stratified. Sea levels are rising, while salinity and micronutrient concentrations are losing uniformity. Moreover, ocean currents have begun to slow down (Berwyn, 2018). The thermohaline cycle involves the cycling of warmer, fresher, and less dense pelagic (surface) water with colder, denser, saltier benthic (deep-sea) water. This allows for the mixing of nutrients, the distribution of heat, and the regulation of climate. Analysis of past climate patterns reveals that a slower thermohaline cycle has been associated with more extreme climate patterns (Berwyn, 2018). However, it is important to note that changes in ocean conditions are not uniform, but rather, vary extensively by region (Winder & Sommer, 2012). That means environmental conditions, which impact the nature of phytoplankton populations, are not homogenous, adding a layer of complexity when determining the impacts they are to face.

Similarly, phytoplankton are undergoing some overarching changes. Common trends include shifting phenology, a change in preferences towards smaller, more buoyant cells, and poleward migration (Ratnarajah et al., 2023). However, under the surface, population modifications are far more complex. For example, certain groups are favored under eutrophic conditions, that is, conditions where there are excessive micronutrients, leading to an unhealthy amount of growth in algal blooms that deplete ecosystem resources, whereas others under fresher or darker conditions (Winder & Sommer, 2012). There are a voluminous amount of environmental factors (e.g., light, heat, nutrients, pH, salinity, etc.) that impact phytoplankton dynamics (Winder & Sommer, 2012). Moreover, each species operates under different sets of ideal conditions. This raises a dilemma. To illustrate this, consider two phytoplankton species living in the same area. Suppose that one species can tolerate a pH range of 5.9 to 6.5, whereas another one tolerates a range of 6.7 to 7.3. With ocean acidity changing heterogeneously, if one area of the ocean has a pH of 6, and another area a pH of 7, then each species would migrate to the area matching their respective preferences, heavily modifying taxonomic composition, biomass, exportation, and other dynamics. However, there are other influential environmental factors, making it important to consider how multiple factors simultaneously impact dynamics. Using the example given, would another factor, such as dissolved oxygen, have precedent over pH when it comes to these species seeking ideal conditions? Moreover, these migrations would leave predators bereft of a major source of food. How would that impact the entire ecosystem? What climatic shifts may result? The circumstances and questions raised by a scenario like this capture the essence of what this study aimed to address.

Examples of Parametric Variability

Parameters that influence phytoplankton conditions are present at the molecular, genomic, cytological, and ecological level. Changes in their values can impact various important biological characteristics, including primary production and metabolic rates. For instance, biochemical processes like DNA methylation, whereby a methyl functional group is applied to the fifth carbon in the carbon ring of the nitrogenous base of cytosine, with warming ocean temperatures, has been found to inhibit amino acid metabolism, as well as respiration and photosynthesis in phytoplankton, while enhancing fatty acid metabolism (Wan et al., 2023). This means that there is a slower rate of primary production and carbon sequestration, inhibiting phytoplankton's role both as the base of marine food chains and as climate regulators. However, seeing as ocean temperature shall change heterogeneously, the extent to which this trend occurs shall vary.

Meanwhile, micronutrients also play a major role in influencing metabolic rates. For example, phosphorus is an integral component of all forms of metabolism, making phosphorus-containing compounds crucial for phytoplankton. However, as discussed above, varying levels of micronutrients, including these compounds, impact dynamics in different ways. It has been found that increased phosphorus levels has allowed for all metabolic processes to occur at faster rates, bolstering the ability of phytoplankton to sequester carbon and provide greater biomass for its predators. However, excessive phosphorus concentrations can be toxic and lead to eutrophication (Li et al., 2023). Moreover, toxicity and metabolic rates vary across different species.

Another example of significant environmental variability is water temperature. Different genera of phytoplankton exhibit different responses to warming ocean temperatures. For example, using a modified Eppley Curve, an exponential function that models the relationship between growth rates and water temperature, one analysis found that, while growth rates are expected to increase alongside temperature, the rate at which the growth rate increases for diatoms was greater than that of dinoflagellates, cyanobacteria, and coccolithophores (Anderson et al., 2023). Additionally, dissimilar thermal attributes are predicted to result in differential migration patterns among different functional groups.

In conjunction with the explanation offered in the previous section, these examples illustrate that for any environmental parameter, there is a great amount of nuance when it comes to the impact that phytoplankton face. This nuance only expands when multiple variables are considered in tandem. It is extremely difficult to perform an experiment that involves multiple independent variables, as confounding factors would easily arise. The alternative would be to perform an experiment using only one variable, which would fail to account for the multifactor interactions that occur. The results of such a procedure across different instances would also vary, failing to paint a solid picture of the impact of that one parameter (Chang et al., 2022).

Computational Modeling of Phytoplankton Dynamics: Progress and Current Limitations

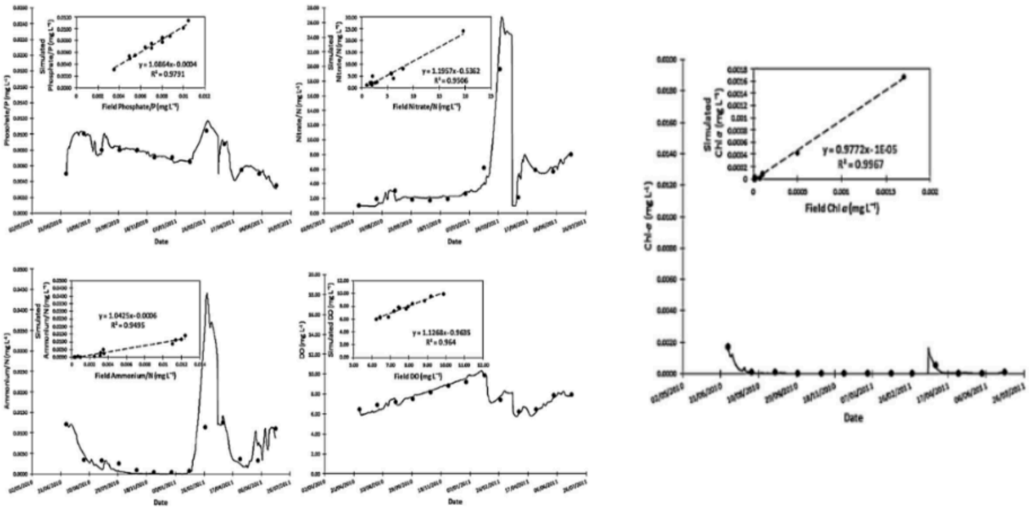
As a result, a computational modeling approach is imperative, as it can be used to capture the nuances of this situation, and provide greater insight into what the observed results signify. In essence, this is what the goal of this project is: to take the complex relationships in phytoplankton populations, and organize, synthesize, and contextualize them, delineating ramifications.

Presently, there are many limitations with computational models of phytoplankton dynamics. One major limitation is the misunderstanding of the role zooplankton play in the modeling process. Different models have made different assumptions about how zooplankton interact in ecological systems, leading to divergent predictions in climate and food web scenarios (Rohr et al., 2023). Indeed, it has been found that more robust data collection methods and raw data on zooplankton is necessary (Ratnarajah et al., 2023). It is a dearth in overall data that limits the predictive power of these computer models. There is a particular lack of data from the Southern hemisphere (Deppeler & Davidson, 2017).

That is not to say that accurate models have not been developed. In fact, there have been models developed for small bodies of water, such as the Tucuruí reservoir in Pará, Brazil (Deus et al., 2013). This computer model was based off of field data on chlorophyll a, dissolved oxygen, ammonia. Through linear regression analysis including R^2, root mean square error, and the slope of regression lines comparing computer predictions to actual results, it was determined that the model was in fact accurate. Figure 1 (Deus et al., 2013) depicts the linear regression between the predicted and field values of these parameters. With extremely high R^2 values, the model was deemed fit to perform other functions within study. This provides a strong example for how the accuracy in computer model predictions can be assessed, allowing for model results and ramifications to be validated. Indeed, validation relies on some form of statistical analysis, which varies from model to model.

Figure 1

An Example of Computational Model Validation Techniques: Tucuruí Reservoir as a Case Study



Note. Each parameter contains a larger graph depicting the raw comparison between field data and computer predictions. From lop left to bottom right, the parameters shown are phosphate, nitrate, ammonia, dissolved oxygen and chlorophyll a. Embedded within are the linear regressions that compare the computer model predictions against the actual field data. Therein lie the R^2 values which serve to evaluate model accuracy. The R2 values for phosphate, nitrate, ammonia, dissolved oxygen, and chlorophyll a were 0.9791, 0.9506, 0.9495, 0.964, and 0.9967, respectively.

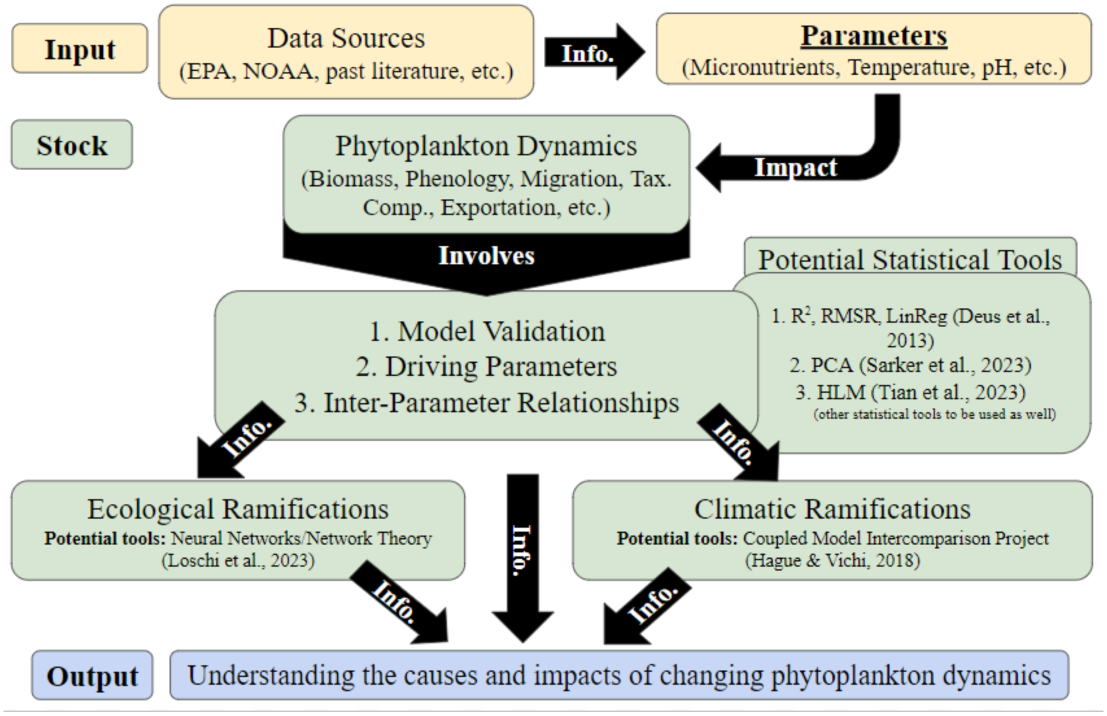
However, different models have been synthesized for different purposes. For example, some models have focused on the identification of driving parameters in phytoplankton dynamics. Using Principal Component

Analysis (PCA), whereby the impact of parameters is measured using vectors, one study of coastal Bangladesh found that salinity, followed by micronutrient concentrations, turbidity, and water temperature played the most significant roles in regulating abundance and spatial variability in phytoplankton (Sarker et al., 2023). Other models have focused on inter-parameter relationships. One study of several lakes in Wuhan, China used a hierarchical linear model. After sorting the parameters into different levels and identifying statistically significant relationships, the one major inter-parameter relationship identified was a negative one between grasslands and water temperature (Tian et al., 2023). From an ecological lens, neural networks have been developed to model the changing flow caused by changing phytoplankton conditions. At a broad level, these networks take in various rates related to energy and matter transfer as parameters, the values of which can be modified to simulate different scenarios. Boit et al. 2012 suggests the gradual implementation of these factors through a series of successive neural networks. When applying this approach to Lake Constance, the fit of the model to predict observed dynamics was maximized, providing a format through which food webs of other systems can be created (Boit et al., 2012). Other studies, such as one of the Venice Lagoon, have been able to identify keystone species (Loschi et al., 2023). From a climatic lens, a focus has been placed on the accuracy of climate models in predicting bloom phenology, as well as other characteristics. The Coupled Model Intercomparison Project (CMIP), with its large scope, has been a particular area of focus. For example, one study found that bloom phenology in the Southern ocean is not accurately predicted as the sea ice concentration levels used in the model were not reflective of on-site levels (Hague & Vichi, 2018). Overall, there exists ample literature describing a myriad of empirical relationships and computational models of the various aspects of the changing characteristics of phytoplankton as well as those ramifications. What is lacking, however, is a unified apparatus to unite these models.

Given the background information and limitations presented, this paper seeked to create a series of computational models bound together as one entire system whereby parametric information on phytoplankton populations could be introduced and results for their populations, and in turn, the environment and climate could be produced. Figure 2 visualizes this overarching computational framework. This study applied this basic framework to data from the National Oceanic and Atmospheric Administration (NOAA)'s comprehensive 2018 World Ocean Database (WOD18). Specifically, the most spatiotemporally cosmopolitan dataset, the Ocean Station Dataset (OSD), was analyzed. These data include millions of casts, spanning multiple centuries and covering virtually the entire ocean (Boyer et al., 2018). Given this impressive scope, this allows the study to take a holistic approach to analysis, partially helping to address the lack of data in computational models. Within the OSD, total oceanic chlorophyll was used as an indicator for primary production. Factors tested include oxygen, micronutrients, pH, salinity, temperature, pressure, and alkalinity. To assess potential forecasting capabilities and overall model strength, a time series of all parameters (including the stated indicator), was created mainly using sinusoidal regression. Subsequently, the relationship of each factor with the indicator was observed using linear regression. Lastly, driving parameters were identified using Principal Component Analysis (PCA).

Figure 2

Proposed Overarching Computational Framework for Modeling of Changing Phytoplankton Dynamics


Input
Data Sources
(EPA, NOAA, past literature, etc.)
Info.
Parameters
(Micronutrients, Temperature, pH, etc.)
Stock
Phytoplankton Dynamics
(Biomass, Phenology, Migration, Tax. Comp., Exportation, etc.)
Impact
Involves
Potential Statistical Tools
1. R^2, RMSR, LinReg (Deus et al., 2013)
2. PCA (Sarker et al., 2023)
3. HLM (Tian et al., 2023)
(other statistical tools to be used as well)
1. Model Validation
2. Driving Parameters
3. Inter-Parameter Relationships
Info.
Info.
Ecological Ramifications
Potential tools: Neural Networks/Network Theory (Loschi et al., 2023)
Info.
Climatic Ramifications
Potential tools: Coupled Model Intercomparison Project (Hague & Vichi, 2018)
Info.
Info.
Output
Understanding the causes and impacts of changing phytoplankton dynamics

Note. This model takes the form of a systems diagram wherein an input is provided for the system stock, operations are performed, and an output is provided. This study proposes that parametric data act as the input, that computational and statistical methods act as the operations within the stock, and that the insights provided on phytoplankton, that is, the study goal, to act as the output. All potential tools proposed above, while useful for achieving their respective ends, however, not all techniques were utilized within this paper.

This apparatus could serve as a viable streamlined process for experts studying phytoplankton populations and their role in the environment and climate. Moreover, it has the potential to serve as a tool for policy makers with regards to water body management. For example, Tian et al. 2023 used results from a multi-agent based model to recommend a controlled increase in micronutrient concentrations and fish that feed exclusively on zooplankton

(Tian et al., 2023). As a whole, this study has provided a potentially potent framework whereby the causes and impacts of phytoplankton conditions can be effectively observed.

Daily Entries:

This section should include specific components to the Engineering Design Process (Build, Test/Evaluate/Revise, Reflection) or Research Process

Abhinav K. Sharma

4 September 2023
Research and Collation of STEM Project Materials
7:15 AM - 10:00 AM; 10:30 AM-1:15 PM; 2:30 PM-5:15 PM

In this work session, previous work from summer assignments were added to the project notes, with extensive revisions to notes (i.e. adding and annotating figures, editing tags, mini-summaries, revising pre-existing notes). Additionally, any missing brainstorming was performed individually (if needed) and pasted into STEM logbook. At the moment, all three ideas have a pie chart, mindmap, and fishbone diagram. Lastly, an area of focus (although not a specific idea) for the STEM project was determined. This idea is regarding studying phytoplankton populations given changing ocean conditions. Much research and brainstorming has been done on this idea already. From this, a timeline for reading some articles was developed (see below). The development of preliminary notes for the Bournedale Elevator Pitch was begun.

Preliminary Notes for Bournedale Elevator Pitch

- Brief Self-Introduction: Hello everyone, I am Abhinav Sharma
- Hook: Our existence is contingent on the well-being of our planet.
- Introduction: This understanding, coupled with my interests in Social and Earth Sciences, has informed my brainstorming for a project idea. With my project, I want to do something that I *know* will help some facet of the environment or climate.
- Goals/"Methods" (i.e. describe brainstorming process): My three pie charts consisted of the topics of food, climatology, and environmental science.
- Results/ "Conclusion" (i.e. how you arrived at idea):
- Plan Moving Forward for Project:
- Clincher:

Timeline for Articles:
By end of day 10 September 2023,
Read the following two articles linked below and complete entries in Project Notes:

1. Monitoring and modelling marine zooplankton in a changing climate | Nature Communications
2. https://link.springer.com/article/10.1007/s10750-022-04795-y?utm_source=getftr&utm_medium=getftr&utm_campaign=getftr_pilot

Abhinav K. Sharma

6 September 2023
Brainstorming and Developing Bournedale Pitch
1:00 PM - 4: 30 PM

During the Bournedale retreat, the 7-Hats brainstorming technique was used to help generate further ideas into our ideas about our STEM project. Additionally, time (both within and without the STEM activity of the retreat) was used to generate my one-minute pitch, which was delivered on 8 September. These activities have helped narrow down two approaches for my project: either using computer modeling to model changes in phytoplankton populations, or performing a laboratory experiment to more deeply investigate the impact of one variable itself.

Abhinav K. Sharma

10 September 2023
Updating Project Notes Document and Research
7:30 AM-12:00 PM; 1:00 PM-3:00 PM; 3:45 PM - 6:30 PM

This session included an overhaul of the project notes file and extensive research. The project notes file was modified in format to be more readable (i.e. notes were transformed into bullet point format). Additionally, previous articles that were read without taking bulleted notes were re-read in order to both add these notes and reinforce understanding of material, which previously was lacking. The “Tags” and “Knowledge Gaps” sections were extensively added to, and the table of contents was updated. The 7-hats brainstorming from Bournedale was also added.

Presently, the most major challenge is developing a precise researchable question/engineering problem/mathematical conjecture. To help achieve this, a three-pronged research plan was adopted, now that general knowledge regarding phytoplankton has been established. One area of focus relates to how phytoplankton dynamics are computationally modeled. Another area of research relates to the impact of specific variables on phytoplankton population. The third area considers how phytoplankton dynamics can be observed through the lenses of different fields of science. Research in all three of these domains has been completed, but there remains other areas to be investigated detailed below. In analyzing the insights of this research, the aim is to develop a series of possible ideas to do for the STEM project, and narrow down an exact project from there.

Another aim is to address the logistics of this project. In order to alleviate any anxiety with regards to time management, a time table of goals has been established below, although they may be subject to change (indeed, this timeline has been revised from the one provided in the previous entry). The first task on the list is to research possible WPI laboratories and facilities for where a STEM project may be conducted. From there, other facilities may be researched. As an exact idea for a project is narrowed down, logistics will be more precisely determined.

Timeline for STEM Project:
Areas of further research to be carried out:

- Deeper focus on computer modeling of phytoplankton
- Analyzing phytoplankton from varying fields of science (ones that have not already been considered- completed ones include Biochemistry, Ecological, Climatology):
    - Astronomy/Exoplanets
- Focusing on oceanic variables that have not yet been researched
    - Salinity
    - Nitrates
    - Iron Fertilization
    - Other trace nutrients
    - Deoxygenation

Task	Due Date
Research WPI Laboratories and Facilities	End of Day 11 September 2023
Finish Notes on Article #6	End of Day 12 September 2023
Finish Notes on Article #7	End of Day 14 September 2023
Finish Notes on Article #8	End of Day 17 September 2023
Complete research on the items listed above. Start developing a list of specific ideas.	End of Day 17 September 2023
Finish Notes on Article #9	End of Day 19 September 2023
Finish Notes on Article #10	End of Day 21 September 2023
Finish Notes on Article #11	End of Day 24 September 2023

Abhinav K. Sharma

11/12 September 2023
STEM Update Google Forms, Pre-Project Planning, Research of WPI Laboratories and Facilities
3:00 PM - 4:30 PM; 9:00 PM - 12:15 AM

During this work session, STEM update google forms #1 and #2 were filled out and submitted. In addition, the pre-project planning document due 13 September was begun. It is to be continued tomorrow. Completing these activities has reinforced my current vision for my project, that is, doing deliberate research and brainstorming to develop an exact idea, and figuring out logistical matters from there. However, a good thing is that it has also forced me to consider other ideas that I have brainstormed in more depth. It must be ensured that they are developed enough such that they can serve as effective ideas to fall back in for times when this project goes wrong. Updated timeline/task list below.

WPI Laboratory Resources:
Using the Schools, Departments, and Programs page on the WPI website as a starting base, possible WPI laboratories and facilities to be used from the STEM project were investigated among different fields. Below are possible facilities for consideration:

- Kaven Hall: Computing, Biology, and Environmental Analysis
    - Possibly the Fuller Laboratory
- Life Sciences and Bioengineering Center: Various forms of chemistry, biology, and biological and chemical engineering
    - There are similar opportunities at Goddard Hall, where this facility is located.
    - This facility contains Vivarium, which houses aquatics for experiments.
- Biological Interaction Forces Laboratory: Focuses on examining biological systems at the nanoscale. This includes use of advanced computer imaging software, bacteria, and liquid environments.

Note that research done by various professors is also listed (link provided below to continue research) The above list merely serves as a snapshot of the sheer amount of resources offered at WPI. Given this abundance, hopefully finding a laboratory for carrying out my experiment for STEM project will not be too difficult a task.
https://www.wpi.edu/academics/schools-departments-programs

Task	Due Date
Finish Notes on Article #6	End of Day 12 September 2023
Complete STEM Pre-Project Planning **(HW)**	End of Day 12 September 2023
Finish Notes on Article #7	End of Day 14 September 2023
Finish Notes on Article #8	End of Day 17 September 2023
Complete research on the items listed above. Start developing a list of specific ideas.	End of Day 17 September 2023
Finish Notes on Article #9	End of Day 19 September 2023
Finish Notes on Article #10	End of Day 21 September 2023
Finish Notes on Article #11	End of Day 24 September 2023

Abhinav K. Sharma

12 September 2023
Pre-Project Planning, Article #6 and #7 Notes
4:45 PM - 6:00 PM; 8:00PM - 11:45 PM

During this session, the pre-project document was completed. Then, notes on articles #6 and #7 were taken on the Project Notes Files. Neither have been finished, but both have been read and have bulleted notes. Below are article resources for my two secondary topics in the event that they end up being my STEM project. Update Tasklist and Timeline Below.

Articles and resources relating to project idea about composting and reforestation:
Mitigation of Greenhouse Gases Emission through Food Waste Composting and Replacement of Chemical Fertiliser
Cost-Benefit and Greenhouse-Gases Mitigation of Food Waste Composting: A Case Study in Malaysia
State of the art and future concept of food waste fermentation to bioenergy - ScienceDirect
Sustainable processing of food waste for production of bio-based products for circular bioeconomy - ScienceDirect
Global warming potential of food waste through the life cycle assessment: An analytical review - ScienceDirect
Quantifying the carbon footprint of household food waste and associated GHGs in Oakville, Ontario, and a municipality's role in reducing both food waste and GHGs - Gooch
Accounting for the Impact of Food Waste on Water Resources and Climate Change

Articles and resources relating to project idea about modeling food waste:
Food Spoilage - an overview | ScienceDirect Topics
Microbial Spoilage of Foods: Fundamentals
Microbiological spoilage of foods and beverages
Managing microbial food spoilage: an overview
FRI BRIEFINGS Microbial Food Spoilage — Losses and Control Strategies
Natural antimicrobial agents to improve foods shelf life
Essential Oils: Sources of Antimicrobials and Food Preservatives
Main Groups of Microorganisms of Relevance for Food Safety and Stability - PMC
Microbial metabolites in nutrition, healthcare and agriculture - PMC
Predictive Modeling of Microbial Behavior in Food - PMC
Handheld DNA sequencers show promise for monitoring microbes during food production

Task	Due Date
Finish Notes on Article #6	End of Day 14 September 2023
Finish Notes on Article #7	End of Day 14 September 2023
Finish Notes on Article #8	End of Day 17 September 2023
Complete research on the items listed above. Start developing a list of specific ideas.	End of Day 17 September 2023
Finish Notes on Article #9	End of Day 19 September 2023
Finish Notes on Article #10	End of Day 21 September 2023
Finish Notes on Article #11	End of Day 24 September 2023

Abhinav K. Sharma

18 September 2023
Research
8:45 AM - 9:45 AM

The time given in class was used to continue research. Notes on Article #6 were taken. Further research was conducted and more articles were found.

Abhinav K. Sharma

20 September 2023
STEM Update Form #3 and Research
10:30 PM - 11:30 PM

During this work session, the third Weekly Update Form was completed. This activity really opened my eyes to the need to hasten up work for the STEM project. Although I have completed some brainstorming and research, there is much that I still have to do, and I have not kept up with the deadlines, indicating that one, better time allocation for STEM is needed, and two, that more realistic goals need to be set. Therefore, the following parameters have been outlined below. After completing the form, some more research was done. Articles to take notes on were added to the Project Notes File.

By the end of September:

- Read all articles that have been posted in the Project Notes File
- Find and take notes on articles on previous knowledge gaps that have been identified (complete these by the end of the month, too)
- Generate and Narrow Down a List of Specific Project Ideas
- Complete Research on WPI Professors and their Research, Research Computational Models of Phytoplankton, and research other pertinents resources

Abhinav K. Sharma

22 September 2023
Research
12:45 PM - 1:45 PM

The time given in class was used to continue research. Notes on Article #6 were continued.

Abhinav K. Sharma

24 September 2023
Research and Project Planning
9:00 PM - 11:30 PM

During this work session, research on the various articles that have been collected was performed. More articles about the knowledge gaps identified above were found. Some of these articles have been posted into Project Notes and are to be read. However, some articles have yet to be pasted into the document, in order to maximize the variety of the content research and avoid repetition. A list of potential ideas was begun and is pasted below.

Areas of further research to be carried out:

- Deeper focus on computer modeling of phytoplankton
    - Findings:
        - Direct input of monitoring data into a mechanistic ecological model as a way to identify the phytoplankton growth-rate response to temperature variations
        - Understanding opposing predictions of Prochlorococcus in a changing climate
        - Three-dimensional model for analysis of spatial and temporal patterns of phytoplankton in Tucuruí reservoir, Pará, Brazil - ScienceDirect
        - A mathematical model of the nutrient dynamics of phytoplankton in a nitrate-limited environment
- ~~Analyzing phytoplankton from varying fields of science (ones that have not already been considered- completed ones include Biochemistry, Ecological, Climatology):~~
    - ~~Astronomy/Exoplanets~~
    - ^^^^No longer interested/beyond the scope^^^^
- Focusing on oceanic variables that have not yet been researched
    - Salinity
        - Findings:
        - Impacts of Temperature, CO2, and Salinity on Phytoplankton Community Composition in the Western Arctic Ocean
        - Environmental Controls of phytoplankton in the river dominated sub-tropical coastal ecosystem of Bangladesh - ScienceDirect
        - Influence of ocean acidification on thermal reaction norms of carbon metabolism in the marine diatom Phaeodactylum tricornutum - ScienceDirect
    - Nitrates, Silicates, & Other trace nutrients
        - Findings:
        - Environmental Controls of phytoplankton in the river dominated sub-tropical coastal ecosystem of Bangladesh - ScienceDirect
        - Effects of Environmental Concentrations of Total Phosphorus on the Plankton Community Structure and Function in a Microcosm Study - PMC
        - Sulfur and phytoplankton: acquisition, metabolism and impact on the environment - Giordano - 2005 - New Phytologist
        - Climate Change Impacts on the Marine Cycling of Biogenic Sulfur: A Review - PMC
    - Iron Fertilization
        - Increasing iron concentration inhibits the Al-incorporation into the diatom biogenic silica: From laboratory simulation of ocean iron fertilization
        - Modelling the ecosystem response to iron fertilization in the subarctic NE Pacific: The influence of grazing, and Si and N cycling on CO2 drawdown

    - ~~Deoxygenation~~
        - Extraneous, other areas of focus have already been found.
- Other findings (mostly about fatty acids):
    - Fatty Acid Profiles and Production in Marine Phytoplankton - PMC
    - Tracking Fatty Acids From Phytoplankton to Jellyfish Polyps Under Different Stress Regimes: A Three Trophic Levels Experiment
    - Lipids of different phytoplankton groups differ in sensitivity to degradation: Implications for carbon export

Potential Ideas:

- Does an increase in ocean temperature decrease phytoplankton contributions to the biogenic sulfur budget?
- Does a decrease in ocean pH decrease the production and upper trophic pyramid-bound distribution of fatty acids in phytoplankton?
- How does phytoplankton cell size impact photosynthetic output?
- ~~Does a decrease in ocean pH decrease metabolic rates in phytoplankton?~~
- Does an increase in the concentration of phosphate and other phosphrous-containing biological compounds increase the rate of carbon, amino and fatty acid metabolism in phytoplankton?

Abhinav K. Sharma

25/26 September 2023
Research
10:45 PM - 1:00 AM

Notes on Article #6 were completed.

Abhinav K. Sharma

26 September 2023
Research and Completion of STEM Update Form #4
8:00 PM - 11:15 PM

Notes on Article #8 were taken, and the STEM Update Form #4 was completed. Overall, time management and diligence with the project has improved. It is necessary to maintain a rapid rate of work so as to keep up with the timeline. Making good progress for meeting the goals of completing research and ideas list by the end of September, although this deadline may be moved back to around October 1-3 in order to ensure that quality research is conducted and that quality work is produced.

Abhinav K. Sharma

1 October 2023
Research for Project Notes and Resources, Professional Communication

7:30 AM - 12:00 PM; 12:45 PM - 2:45 PM; 6:15 PM - 8:15 PM; 10:00 PM - 11:30 PM

During this work session, extensive work was carried out on Project Notes document. Entries for articles #7 and #8 were completed. Despite the fact that there are still a lot more articles to be read, many good ideas have been brainstormed, providing a solid base to go off of for a project idea. Rapid, diligent work, however, is still of the essence. Additionally, WPI laboratory resources were investigated more in depth. It is possible that some of the professors listed below may be contacted soon. Professional communication for Article #8 was written. The professional communication done for article #7 previously is also pasted below.

Potential Ideas:

- Does an increase in ocean temperature decrease phytoplankton contributions to the biogenic sulfur budget?
- ~~How does phytoplankton cell size impact photosynthetic output?~~
- Does a decrease in ocean pH decrease the production and upper trophic pyramid-bound distribution of fatty acids in phytoplankton?
- ~~Does a decrease in ocean pH decrease metabolic rates in phytoplankton?~~
- Does an increase in the concentration of phosphate and other phosphrous-containing biological compounds increase the rate of carbon, amino and fatty acid metabolism in phytoplankton?
- How does a decrease in carbon metabolism in phytoplankton impact to transfer of crucial fatty acids up the trophic pyramid?
- How does iron fertilization differentially impact the rates of catabolism and anabolism of amino acids? How does this change given different concentrations of iron?
- How do diversions from the Redfield Ratio impact the range of light phytoplankton are able to absorb and use as energy during photosynthesis?

Emails:
Article #7:
To: L.Ratnarajah@liverpool.ac.uk
Lead-Author of “Monitoring and modelling marine zooplankton in a changing climate”
https://www.nature.com/articles/s41467-023-36241-5.pdf

Subject Line: Questions Regarding Recent Publication (Monitoring and modelling marine zooplankton in a changing climate)

Email Body:

Greetings, Dr. Ratnarajah.

I am Abhinav K. Sharma, a student at the Massachusetts Academy of Math and Science (MAMS) at the Worcester Polytechnic Institute (WPI). I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. In conducting research, I have read your article entitled

“Monitoring and modelling marine zooplankton in a changing climate” from Nature Communications published earlier this year.

I have found that the contents of the article were incredibly helpful in developing my project. Content wise, it was easy to understand both the overall trends zooplankton are undergoing, and the necessary future steps needed for sampling and researching them. The insights offerred have helped me make connections to my project regarding phytoplankton. Specifically, my project relates to modeling the impact of changing ocean conditions on phytoplankton at a global scale, taking into account regional variations, and various biotic and abiotic conditions.
That being said, I wanted to ask a few questions regarding your article:

A major problem cited in your article was the lack of resources to model zooplankton dynamics and the factors that cause that. Do similar problems exist when it comes to modeling phytoplankton?

How are phytoplankton impacted by the trends in zooplankton populations outlined in your article? Indeed, many of the findings in your article about zooplankton have similarities to what I have found in my research regarding phytoplankton.

As such, in researching zooplankton populations, I hope to make important connections to phytoplankton as I work on my project. If you could offer any additional insights, that would be much appreciated. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

(Article #8)
To: pengjin@gzhu.edu.cn
Lead-Author of “DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change”

Subject Line: Questions and Interest Regarding Recent Publication (DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change)

Email Body:

Greetings, Dr. Jin.

I am Abhinav K. Sharma, a student at the Massachusetts Academy of Math and Science (MAMS) at the Worcester Polytechnic Institute (WPI). I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project relates to investigating the impact of how changing ocean conditions are impacting phytoplankton. In conducting research, I have read your article entitled “DNA methylation and gene transcription act cooperatively in driving the adaptation of a marine diatom to global change” published earlier this year in the Journal of Experimental Botany.

I have found that the contents of the article were incredibly helpful in developing my project. It has provided important insights about the genetic and biochemical dynamics that phytoplankton populations are facing. I find the possibility of DNA methylation being a possible means of adaptation to be truly fascinating. These ideas have been helpful in developing my project regarding phytoplankton.

That being said, I wanted to ask a few questions regarding your article:

Beyond temperature and CO_2 concentration, are there other variables (for example, dissolved oxygen or salinity), that impact gene regulation, and thereby metabolism in phytoplankton?

Are there other biochemical adaptations beyond DNA methylation that phytoplankton can use? Do these methods have any connection to DNA methylation?

How do diversions from the Redfield Ratio impact the rate of DNA methylation in phytoplankton? How does this impact metabolic rates?

If you could offer any additional insights, that would be much appreciated. Additionally, if you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

WPI Laboratory Resources:
Reeta Prusty Rao| Worcester Polytechnic Institute (Department Head of Biology and Biotechnology)
Luis Vidali| Worcester Polytechnic Institute (Focus: plants)
Anita Elaine Mattson| Worcester Polytechnic Institute (Department Head of Chemistry and Biochemistry)
Arne Gericke| Worcester Polytechnic Institute (Focus on protein/lipid systems could possibly be connected to a phytoplankton project idea)
~~~~~~

Using the Schools, Departments, and Programs page on the WPI website as a starting base, possible WPI laboratories and facilities to be used from the STEM project were investigated among different fields. Below are possible facilities for consideration:

- Kaven Hall: Computing, Biology, and Environmental Analysis
    - Possibly the Fuller Laboratory
- Life Sciences and Bioengineering Center: Various forms of chemistry, biology, and biological and chemical engineering
    - There are similar opportunities at Goddard Hall, where this facility is located.
    - This facility contains Vivarium, which houses aquatics for experiments.
- Biological Interaction Forces Laboratory: Focuses on examining biological systems at the nanoscale. This includes use of advanced computer imaging software, bacteria, and liquid environments.
~~~~~~

Questions for First Update Meeting:

- Any suggestions for laboratory resources?
- How do I know when I can send emails?

Abhinav K. Sharma

2 October 2023
Research for Project Notes
1:45 PM - 2:30 PM

Class time was used to take notes on Article #9 in the Project Notes (before having STEM Update Meeting #1).

Task items from STEM Update Meeting #1:

- Contact laboratory resources
- Look into that phosphate idea more, micronutrient stuff. Probably eliminate computer modeling at this point as a project idea, but read articles as the techniques may still be relevant.

Abhinav K. Sharma

9 October 2023
Research for Project Notes
5:00 PM - 7:00 PM; 9:00 PM - 11:30 PM

Emails to Arne Gerike and Luis Vidali were drafted and sent. Emails have been pasted below (and have now been pasted under the Professional Communications section). Additionally, Project Logbook was updated to include research questions and hypotheses. Research notes on Article #9 were begun.

Luis Vidali:
Subject Line: Questions Regarding Research from a Mass Academy Student

Email Body:
Greetings, Dr. Vidali.

I am Abhinav K. Sharma, a student from Mass Academy at WPI. I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project aims to investigate the impact of global warming-induced oceanic changes on phytoplankton at a biochemical level. As part of my project development, I have investigated some of your work due to its focus on autotrophic organisms.

I notice that your research centers around the cellular dynamics of plants with a particular focus on the cytoskeleton. With significant experience with the cytology and biochemistry of autotrophic organisms, I was wondering if you could answer some of the following questions:

How applicable are the findings of the model organism *Physcomitrella patens* to marine autotrophs, including phytoplankton? Is it too taxonomically separate from them for findings to be extrapolated?

Among the autotrophs that you have worked with, how do different rates of cellular growth, communication, and transportation brought about by varying conditions in the cytoskeleton impact metabolic and photosynthetic rates?

If you could offer any additional insights, that would be much appreciated. Additionally, I am interested in learning more about your research in depth. If you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

Arne Gerike (works with proteins and lipids):
Subject Line: Questions Regarding Research from a Mass Academy Student

Email Body:
Greetings, Dr. Gerike.

I am Abhinav K. Sharma, a student from Mass Academy. I am currently working on developing a five-to-six-month research project regarding phytoplankton for a STEM science fair. Specifically, my project aims to investigate the impact of global warming-induced oceanic changes on phytoplankton at a biochemical level. In particular, I have considered investigating how varying concentrations of phosphates and other phosphorus-containing compounds impact metabolic rates in phytoplankton. As part of my project development, I have investigated some of your work due to its focus on interactions between proteins and lipids.

In particular, I notice that you have a particular focus on phosphoinositides and their various interactions with proteins. With significant experience with phosphorus-containing molecules and biochemical interactions, I was wondering if you could answer some of the following questions:

What factors control the type and amount of phosphoinositide lipids found in a cell?

How does phosphoinositide composition and concentration differ across different organisms?

Similarly, how do protein interactions with phosphoinositide lipids differ among terrestrial and aquatic organisms? Among autotrophic and heterotrophic organisms?

If you could offer any other insights, that would be much appreciated. Additionally, I am interested in learning more about your research in depth. If you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

Abhinav K. Sharma

10 October 2023
Research for Project Notes, STEM Update Form
2:00 AM - 5:45 AM; 7:15 AM - 7:45 AM; 9:45 AM - 10:15 AM; 3:45 PM - 4:30 PM

Research notes on Article #9 were completed, and final preparations for STEM Update meeting #2 were made (pasted below). In addition, STEM update form #5 was completed.

Updates:

- Reached out to Arne Gericke, Luis Vidali + 2 other authors (both are in foreign nations, though)
- Developed a timeline for my project
    - End of October Break: Develop *exact* project idea, finish research through article #15
    - B term: Complete entire procedure, proactively complete assignments (patents/20 articles by mid-november)
    - C Term: Do data analysis, develop STEM fair products, other assignments
- Content-wise, heading towards a more lab-based project, but I might implement AI/ computational techniques in some manner.
- Been investigating the more biochemistry side of things

Questions:
For one of the authors I tried emailing, the email did not go through. I tried again through various email accounts, but there too I was unable to send my email. What do you recommend for such situations?

Article:
I read an article from the American Society of Microbiology's Journal Microbiology Spectrum entitled "Two-sided effects of the organic phosphorus phytate on a globally important marine coccolithophorid phytoplankton" published just last month by P.I. Senjie Lin of UConn University.

For a brief summary:
This study aimed to establish a clearer relationship between varying concentration of phytic acid, a cosmopolitan phosphorus-containing compound and the physiological characteristics of the also cosmopolitan coccolithophore *Emiliania huxleyi*, which up to this point has not been well understood. Four treatments, including a control, phytic acid, dissolved inorganic phosphorus, and a combined treatment of those two were biochemically analyzed for cell size and concentration, the amount of lipids, photosynthetic output, and nutrient stoichiometry. Gene expression was also analyzed by extracting RNA, converting it into DNA via reverse transcriptional and purification methods, and comparing that DNA to the *E. huxleyi* to identify DEGs. Findings show that phytic acid upregulated genes relating to carbohydrate, lipid and amino acid metabolism and cell signaling and membrane development, and that, contrary to popular belief in the field, dissolved inorganic phosphorus is not preferentially absorbed over

phytic acid. However, an area that still needs further investigation is the potential toxic effects phytic acid may have on phytoplankton as a whole.

Look at pubmed to see where authors have moved on if you can no longer find them.

Contact senjie lin → long distance contribution to their lab? See if you can do long distance
Analysis of chemical data, gene expression
Develop a backup plan!
Get genes and then do the simulation. Create a predictive model. Extrapolate to other micronutrients → good backup ideas! Pitch to this P.I. Utilize data (with his mentorship) → create a model
Predicting algal blooms

Net Logo → good tool! Lots of guides to create simulations. WPI campus resource: Elizabeth Ryder, bee populations → for algal blooms of phytoplankton (also look into the literature too, and then how those were set up, etc.)

Contributions to labs → good opportunity but change idea

So much stuff to consider!

Abhinav K. Sharma

12 October 2023
Project Notes
7:45 AM - 8:45 AM;
Given class time was used to update logbook (adding captions to the brainstorming items), and continue notes on Article #10 in the Project Notes file.

Abhinav K. Sharma

15 October 2023
Project Notes
10:45 AM - 12:00 PM; 12:45 PM - 2:45 PM; 4:00 PM - 5:30 PM; 6:30 PM - 9:00 PM; 10:30 - 11:15 PM

These work sessions were dedicated to continuing work on Article #10 Notes.

Abhinav K. Sharma

16 October 2023
Fulfillment of A Term Requirements
2:30 AM - 5:45 AM; 6:15 AM - 10:45 AM; 1:15 - 2:45 PM

Work sessions on this day were dedicated to completing article #10 in Project Notes to be in accordance with the ten article requirement. The literature research parameters and tags section were partially updated and shall be more fully updated at a later time. Brainstorming materials were elaborated upon so as to enrich the content of the Project Logbook and provide a detailed narrative of the initial steps of progression for this project. A similar process has been performed for emails in the logbook. This time was also dedicated to making a slideshow, preparing for, and taking notes from the insights of STEM Update meeting #3. Following this meeting, I have a fairly definite resolution to pursue computationally modeling global migratory patterns in phytoplankton given global warming-induced changes in oceanic conditions, and finding what ecological and climatic ramifications may result. This has set a good marker for where I ought to go from here with regards to my project. I will take much of October break to get ahead of assignments and look into NetLogo. I am eager for this journey, albeit apprehensive.

Subject Line: Questions and Interest Regarding Recent Publication (Two-sided effects of the organic phosphorus phytate on a globally important marine coccolithophorid phytoplankton)
Email to Dr. Lin:

I am Abhinav K. Sharma, a student at the Massachusetts Academy of Math and Science (MAMS) at the Worcester Polytechnic Institute (WPI). I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project relates to investigating the impact of how changing ocean conditions are impacting phytoplankton at a biochemical level. In particular, I have considered investigating how varying concentrations of phosphates and other phosphorus-containing compounds impact metabolic rates in phytoplankton. In conducting research, I have read your article entitled “Two-sided effects of the organic phosphorus phytate on a globally important marine coccolithophorid phytoplankton” published just last month in the Microbiology Spectrum Journal.

I have found that the contents of the article were incredibly helpful in developing my project. It has provided important insights regarding the role phytic acid plays in upregulating many metabolic processes, but also its possible toxic effects. These ideas have been helpful in developing my project regarding phytoplankton. I also notice that you are from UConn, and work a lot with the ecology and biochemistry of phytoplankton and marine environments.

That being said, I wanted to ask a few questions regarding your article and work:

How do the varying cytological attributes of different phytoplankton species influence the optimum phosphate levels under which they can operate?

Does an enhancement in metabolic processes among phytoplankton result in an enhancement in metabolic processes among other organisms higher up the trophic pyramid? How do varying metabolic rates impact symbiotic activity in phytoplankton?

Have any of your PhD students worked remotely with laboratory-collected data, specifically DNA or molecular data?

If you could offer any additional insights, that would be much appreciated. Additionally, if you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

Email to Dr. Ryder:
Greetings, Dr. Ryder.

I am Abhinav K. Sharma, a student from Mass Academy here at WPI. I am currently working on developing a five-to-six-month research project regarding phytoplankton for a science fair. Specifically, my project aims to investigate the impact of global warming-induced oceanic changes on phytoplankton. In doing so, I plan to utilize computer modeling. As part of my project development, I have investigated some of your work due to its implementation of computational techniques.

I notice that your work has involved using NetLogo to create variables that were then used to model bees and their pollination patterns, and that your class, Simulation in Biology, involves students building their own simulations of biological systems. Given this experience, I was wondering if you could answer some of the following questions:

For someone who has only a very basic background in computer science, how steep of a learning curve does NetLogo pose?

Do you believe it is possible to create predictive ecological and climatic models of oceanic systems at a global scale, taking into account the impact of global warming-induced ocean changes on phytoplankton? What level of complexity can programs on NetLogo be?

Have any of your past students worked in any area that relates to what I mentioned in the previous question?
If you could offer any additional insights, that would be much appreciated. Additionally, I am interested in learning more in-depth about your work. If you have the time, I would appreciate an opportunity to interview you as part of my project development. Thank you very much for your time and expertise.

Regards,
Abhinav K. Sharma

Updates for STEM meeting #3:

- Reached out to both Senjie Lin and Elizabeth Ryder
- Dr. Gericke and Dr. Vidali have yet to get back to me.

I see my project going either one of two ways:

- Biochemical lab analysis of micronutrients, if not, otherwise:
- Computational modeling of micronutrients (computational modeling using NetLogo and other software)

        - Algal bloom prediction
        - Modeling migration patterns and shifting ecologies and food webs
            - Predicting climatic/ecological effects based off that

For October Break, my plan is to:

- Be proactive with the long term assignments
    - 20 articles, 3 patents, MSEF proposal
- Look into NetLogo (learning curve?)
- Developing various testing plans / experimental designs

If I don't know what approach I ought to take, how can I draft my MSEF proposal? How would you recommend I proceed in those scenarios?

[present article]

Feedback on slideshow/presentation?

Revise, revise, revise!
It's okay to ask for help. Evaluation of different models. Mention your lack of experience.
Demonstrations, trainings, very good! (Do a bunch of build-something projects basically)
Plenty of time and resources, CS is not impossible!

Data attainment (EPA, MADEP) → model development → Model validation (how?)
You know your stuff, so reduce text on slides.
Very promising, good applicatory/implicatory idea for the project.

Abhinav K. Sharma

21 October 2023
Notes
9:45 AM - 12:00 PM; 12:45 PM - 5:30 PM

Article Notes #11 were worked on. Some adjustments were made with regards to which articles are to be read.

Abhinav K. Sharma

22 October 2023
Notes and Downloading NetLogo
1:15 PM - 3:00 PM; 4:00 PM - 5:00 PM; 6:00 PM - 9:30 PM

Notes on Article #11 were completed. NetLogo was downloaded and installed. Some very preliminary work was done to get acquainted with the software. Other softwares are to be researched.

Abhinav K. Sharma

31 October 2023
Project Notes and MSEF Proposal
12:30 PM - 1:15 PM; 6:00 - 6:45 PM

Work sessions on this date were dedicated to working on the MSEF proposal and reading and taking notes on Article #12.

Abhinav K. Sharma

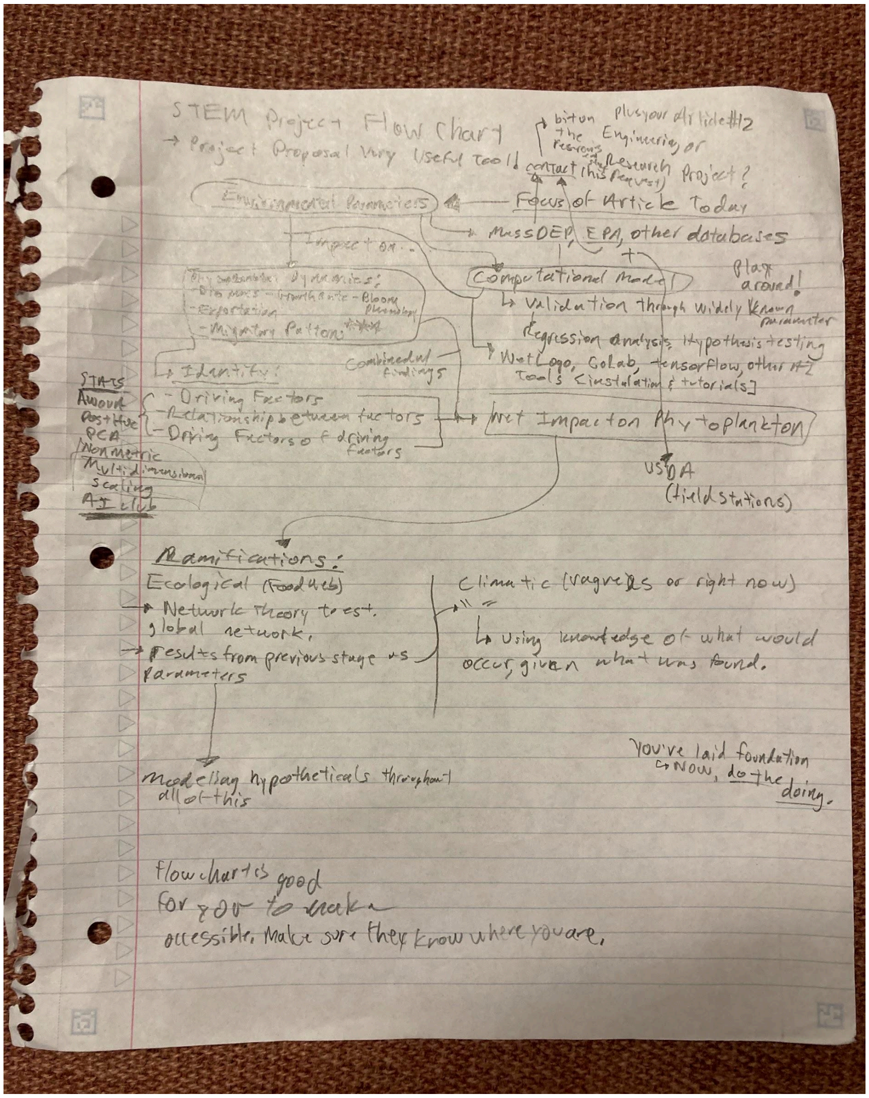
1 November 2023
MSEF Proposal Draft Completion and Project Notes
2:00 AM - 6:45 AM; 9:45 AM - 10:15 AM; 6:00 - 6:45 PM; 8:00 PM - 10:00 PM

Work sessions on this date were dedicated to working on the MSEF proposal and reading and taking notes on Article #12. The first draft of the MSEF proposal was completed on this date.

Abhinav K. Sharma

2 November 2023
Article Notes and STEM Update Meeting Preparation
12:00 AM - 6:45 AM; 4:10 PM - 5:00 PM;

On this day, notes for Article #12 were completed. Additionally, preparations were made for the fourth STEM update meeting. Notably, a flowchart for the course of action for the project was developed and is pasted below. During the update meeting, many courses of action, such as emailing Massachusetts state and federal government officials in the USDA, NOAA, MassDEP, EPA and other relevant organizations, and exploring and experimenting with computational software were determined.


STEM Project Flow Chart
→ Project Proposal Very Useful Tool!
bit on the Previous
Plus your Article #12
Engineering or Research Project?
contact (this Request)
Environmental Parameters
Focus of Article Today
Impact on..
MassDEP, EPA, other databases
+
Phytoplankton Dynamics;
- Bio mass - Growth Rate - Bloom Phenology
- Exportation
- Migratory Pattons
Computational Model
Play around!
↳ Validation through widely known parameter
Regression analysis, Hypothesis testing
WetLogo, Colab, tensorflow, other AI Tools [instalation & tutorials]
Combined all findings
↳ Identify!
- Driving Factors
- Relationship between factors
- Driving Factors of driving factors
Net Impact on Phytoplankton
STATS
Anova
Post Hoc
PCA
Non Metric Multidimensional Scaling
AI club
USDA (field stations)
Ramifications:
Ecological (Food Web)
→ Network theory to est. global network.
→ results from previous stage as parameters
Climatic (vagreies or right now)
" "
↳ Using knowledge of what would occur, given what was found.
modelling hypotheticals throughout all of this
You've laid foundation
↳ Now, do the doing.
flow chart is good for you to make accessible. Make sure they know where you are.

Abhinav K. Sharma 8 November 2023 10:30 AM

Abhinav K. Sharma

6 November 2023
Research and Project Notes
12:45 PM - 1:45 PM
The provided class time was used to continue research and continue project notes. More articles were found, and previous entries were modified and reviewed to ensure understanding of all relevant concepts hitherto.

Abhinav K. Sharma

8 November 2023
Research and Project Notes

3:45 AM - 5:45 AM;
More research was performed, and many more articles were found. Article #13 notes were begun. All research at this point has pertained to the hard mechanics of computational modeling, and collection of the impact of parameters, stratifying for phytoplankton species, and geographic location. The aim is to find data in studies that can be fed into the computer model or models that are to be developed.
Research parameters to investigate:
- Means through which model validation occurs (specifically for phytoplankton dynamics)
- Computational Techniques, Statistical Tools and other assets as keywords in combination with "phytoplankton"

Abhinav K. Sharma

13 November 2023
Project Notes, MSEF Proposal, Initial Investigation of Data Sources
6:15 AM - 7:00 AM; 8:00 - 8:45 AM; 7:15 PM - 8:00 PM

Research on Article Notes #13 was performed. Additionally, the search for potential resources for data to feed into a computational model was begun. The relevant links attained are pasted below:
https://www.epa.gov/caddis
https://www.ncei.noaa.gov/products/world-ocean-database
https://www.mass.gov/guides/water-quality-monitoring-program-data#-data-files-

Potential Data from Article #12 (The Article Utilizing HLM in Wuhan, China)
(Kumar, 2018)
https://www.sciencedirect.com/science/article/pii/S0078323421000981; → Suspended Solids
https://www.sciencedirect.com/science/article/pii/S0278434317306477 → Micronutrients

Abhinav K. Sharma

14 November 2023
Project Notes
6:15 PM - 7:00 PM

Work on completing Article #13 Notes in Project Notes was continued.

Abhinav K. Sharma

15 November 2023
Project Notes
1:45 PM - 2:45 PM; 7:00 PM - 7:45 PM

The given class time and other work session were used to continue Article #13 Notes in Project Notes.

Abhinav K. Sharma

17 November 2023
Project Notes
8:45 AM - 9:45 AM

The given class time was used to continue completing Article #13 Notes in Project Notes.

Abhinav K. Sharma

20 November 2023
Investigation of Data Sources, Project Notes, and Update Form #7
2:00 AM - 4:30 AM; 6:15 AM - 7:45 AM

Research on Article Notes #13 was continued. Additionally, the search for potential resources for data to feed into a computational model was continued. Update Form #7 was also completed. The relevant links attained are pasted below:
https://www.ncei.noaa.gov/products/world-ocean-database (Data from the NOAA)
https://www.mass.gov/guides/water-quality-monitoring-program-data#-data-files- (Data from MassDEP)
https://www.usda.gov/content/usda-open-data-catalog (Data from the USDA)

Sources and Data from the EPA:
https://www.epa.gov/caddis
Provides a broad overview of the scientific process scientists use for testing water quality. It provides in-depth information on stressors and statistical tools, and also provides examples and case studies.

https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys
This link contains specific csv files collected as part of the National Aquatic Resource Surveys (NARS) on phytoplankton dynamics, as well as a myriad of environmental parameters.

Phytoplankton water quality relationships in U.S. lakes: The common phytoplankton genera from eastern and southeastern lakes (note that this source is dated, being from the 1970s, so tread with caution)
Contains the ideal levels for numerous environmental variables for many genera of phytoplankton found throughout the northeastern and southeastern US.

Literature Review on Nutrient-Related Rates, Constants, and Kinetics Formulations in Surface Water Quality Modeling
Provides a series of mathematical models and equations associated with environmental parameters that impact water quality.

People to Contact:
Darryl J. Keith Ph.D. - Biological Oceanographer who has worked on cyanobacteria collection and management in the New England Area
Authors from 2019 Parameter Dataset Linked Above

- Ben Cope (Environmental Engineer)
- Taimur Shaikh (Involved more in the energy/policy field)
- Rabjir Parmar (Project Officer, Computer Scientist)
- Dr. Steven Chapra (Environmental Engineer, Extensive Experience with Computer Modelling for Water Quality)
- Dr. James L. Martin (Mississippi State University Professor, Experience with Environmental Software and Water Quality Modelling)

Potential Data from Article #12 (The Article Utilizing HLM in Wuhan, China)
(Kumar, 2018)
https://www.sciencedirect.com/science/article/pii/S0078323421000981; → Suspended Solids
https://www.sciencedirect.com/science/article/pii/S0278434317306477 → Micronutrients

Kaggle Datasets:
https://www.kaggle.com/datasets/sohier/calcofi
This dataset contains information about not just phytoplankton and stressors; in fact, the main focus is on fish. Ideally, planktonic data can be extracted, in conjunction with
https://www.kaggle.com/code/ankitachoudhury01/water-potability-notebook-seaborn
Although not likely to be used directly for the project (dataset is more about water quality), this is a great tool in that it would act as an example for the necessary Python syntax and statistical tools.
https://www.kaggle.com/datasets/brsdincer/ocean-data-climate-change-nasa/data and
https://www.kaggle.com/datasets/joebeachcapital/global-earth-temperatures
General oceanographic data.

Possible AI/Computational Tools to Use for Model Development:

- NetLogo
- TensorFlow
- Google CoLab
- Scikit-learn
- PyTorch
- Keras
- GitHub (url contains list of source code to build off of)
- CBIOMES
    - This is not an exact software, rather, it is an ongoing project that compiles relevant phytoplankton data to create numerous models crucial for biogeochemical cycling
- FlowCam (seems to be more on the image analysis side but includes time series)
- Pacific Northwest National Laboratory
    - Similar to CBIOMES, this is a publicly available, developing model for biogeochemical cycling that contains multiple parameters, phytoplankton included.

Abhinav K. Sharma

21 November 2023

Pitch Development and Project Notes
6:15 AM - 7:45 AM ; 5:15 PM - 6:30 PM

Preliminary elevator pitches for this project were drafted. Additionally, Article #13 were continued.

Abhinav K. Sharma

24 November 2023
Pitch Development
1:00 PM - 3:15 PM; 8:00 PM - 11:30 PM

Having received substantial peer feedback, and with a fresh mind, both pitches were fully developed and refined. After heavy rehearsal (without audience), the pitches were presented. All parts of the assignment associated with speaking to nontechnical/technical relatives were completed.

Abhinav K. Sharma

26 November 2023
Grant Proposal Checkpoint #2
11:30 AM - 1:00 PM;

The Grant Proposal Checkpoint #2 assignments as part of this larger project were completed.

Abhinav K. Sharma

27 November 2023
Project Notes
6:15 AM - 7:30 AM; 7:15 PM - 7:45 PM

Notes on Article #13 were continued.

Abhinav K. Sharma

28 November 2023
Project Notes
6:15 AM - 7:00 AM; 6:30 PM - 7:00 PM

Notes on Article #13 were continued.

Abhinav K. Sharma

1 December 2023
Research in Conjunction With Grant Proposal Draft #1
12:00 AM - 7:30 AM

During this work session, the grant proposal draft number #1 assignment was completed. Part of this process involved the collation of articles to be read, as well as a considerable amount of possible data sources and computational techniques and software.

Abhinav K. Sharma

5 December 2023
6:30 PM - 7:00 PM

Project Notes on Article #13 were continued.

Abhinav K. Sharma

6 December 2023
2:00 AM - 3:30 AM; 6:30 PM - 7:45 PM

Work on Experiment #1 has begun. A dataset provided by the EPA was imported into a Google spreadsheet, and the data was rearranged. This spreadsheet can be found below.

Abhinav K. Sharma

7 December 2023
1:15 AM - 5:00 AM; 12:30 PM - 1:30 PM; 4:45 PM - 6:30 PM

The process of experimental collection was continued. This part of the process mainly involved the use of this software to carry out one-way ANOVA tests and Post-Hoc Tukey Tests. This was done with the ends of determining the homogeneity of preferences of the environmental parameters among the phytoplankton genera listed in the set. With this information, basic decisions could be made about which parameters to implement for initial model development.

Abhinav K. Sharma

9 December 2023
8:00 AM - 12:00 PM; 12:30 PM - 10:15 PM

Work on this day was dedicated to the creation of the December Fair poster. All necessary tools to this end were collated. This included graphics of the methodology and background, as well as images of the various figures and tables used for presenting data.

Abhinav K. Sharma

11 December 2023
2:00 AM - 5:45 AM

This work session was dedicated to finalizing material for the December Fair, as well as analyzing future data sources and assessing ways in which the present model apparatus could be iterated upon.

Data from the NOAA
https://www.ncei.noaa.gov/products/world-ocean-database

Data from MassDEP
https://www.mass.gov/guides/water-quality-monitoring-program-data#-data-files-

Data from the USDA
https://www.usda.gov/content/usda-open-data-catalog

Sources and Data from the EPA:
https://www.epa.gov/caddis
Provides a broad overview of the scientific process scientists use for testing water quality. It provides in-depth information on stressors and statistical tools, and also provides examples and case studies.

https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys
This link contains specific csv files collected as part of the National Aquatic Resource Surveys (NARS) on phytoplankton dynamics, as well as a myriad of environmental parameters.

Phytoplankton water quality relationships in U.S. lakes: The common phytoplankton genera from eastern and southeastern lakes (note that this source is dated, being from the 1970s, so tread with caution)
Contains the ideal levels for numerous environmental variables for many genera of phytoplankton found throughout the northeastern and southeastern US.

Literature Review on Nutrient-Related Rates, Constants, and Kinetics Formulations in Surface Water Quality Modeling
Provides a series of mathematical models and equations associated with environmental parameters that impact water quality.

https://www.epa.gov/nutrientpollution/nutrient-data
Provides data on pollutants, specifically nitrogen and phosphorus.

Kaggle Datasets:
https://www.kaggle.com/datasets/sohier/calcofi
This dataset contains information about not just phytoplankton and stressors; in fact, the main focus is on fish. Ideally, planktonic data can be extracted, in conjunction with data on fish.
https://www.kaggle.com/code/ankitachoudhury01/water-potability-notebook-seaborn
Although not likely to be used directly for the project (dataset is more about water quality), this is a great tool in that it would act as an example for the necessary Python syntax and statistical tools.
https://www.kaggle.com/datasets/brsdincer/ocean-data-climate-change-nasa/data and
https://www.kaggle.com/datasets/joebeachcapital/global-earth-temperatures provide

General oceanographic data.

Potential Data from Article #12 (The Article Utilizing HLM in Wuhan, China)
(Kumar, 2018)
https://www.sciencedirect.com/science/article/pii/S0078323421000981; Suspended Solids
https://www.sciencedirect.com/science/article/pii/S0278434317306477; Micronutrients

(Li et al., 2020)
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231357; Holistic Analysis

(Wang et al., 2020)
https://doi.org/10.1016/j.marpolbul.2020.111010; Holistic Analysis + Turbidity, Ammonia Nitrogen, Nitrate Nitrogen

(Jakobsen et al., 2015)
https://www.sciencedirect.com/science/article/pii/S0022098115300083; Light, pH, salinity

(Ke et al., 2019)
https://www.sciencedirect.com/science/article/pii/S1872203217301828; Holistic Analysis + Suspended Solids

(Wu et al., 2019)
https://www.sciencedirect.com/science/article/pii/S1470160X19306685; Transparency and Flow Rate

(Jiang et al., 2014)
https://www.sciencedirect.com/science/article/pii/S1470160X14000089?via%3Dihub; Water Temperature

(Cao et al., 2016)
https://www.sciencedirect.com/science/article/pii/S1470160X15005117?via%3Dihub; Chemical Oxygen Demand, Total Nitrogen

(Silva et al., 2014)
https://www.sciencedirect.com/science/article/pii/S0075951114000280?via%3Dihub; Broader ecological scale

Possible AI/Computational Tools to Use for Model Development:

- NetLogo
- TensorFlow
- Google CoLab
- Scikit-learn
- PyTorch
- Keras
- GitHub (url contains list of source code to build off of)
- CBIOMES

    - This is not an exact software, rather, it is an ongoing project that compiles relevant phytoplankton data to create numerous models crucial for biogeochemical cycling
- FlowCam (seems to be more on the image analysis side but includes time series)
- Pacific Northwest National Laboratory
    - Similar to CBIOMES, this is a publicly available, developing model for biogeochemical cycling that contains multiple parameters, phytoplankton included.

People to Contact:
Darryl J. Keith Ph.D. - Biological Oceanographer who has worked on cyanobacteria collection and management in the New England Area
Authors from 2019 Parameter Dataset Linked Above

- Ben Cope (Environmental Engineer)
- Taimur Shaikh (Involved more in the energy/policy field)
- Rabjir Parmar (Project Officer, Computer Scientist)
- Dr. Steven Chapra (Environmental Engineer, Extensive Experience with Computer Modelling for Water Quality)
- Dr. James L. Martin (Mississippi State University Professor, Experience with Environmental Software and Water Quality Modelling)

Additional Resources:
https://bg.copernicus.org/articles/12/4447/2015/
https://www.researchgate.net/publication/304185998_Reviews_and_syntheses_Parameter_identification_in_marine_planktonic_ecosystem_modelling

Abhinav K. Sharma

12 December 2023
8:00 AM - 11:00 AM

During this timeframe, presentations of the project in its current state, which mainly included to the material described in Experiment #1 (see below) were performed for the December Fair. Based on the feedback provided, the preliminary data analysis, as well as the systems diagram of the methodology has established a solid foundation for carrying out the process of the project itself. However, the verbal and presentational aspects need improvement.

Abhinav K. Sharma

13 December 2023
12:00 AM - 6:00 AM

Project notes on article #13 were completed. Moreover, the notes on articles #14-20 were all begun.

Abhinav K. Sharma

14 December 2023
12:00 AM - 6:00 AM; 11:00 AM - 11:30 AM

Project notes on articles #14-20 were all continued. In addition, STEM Update meeting #6 occurred. This mainly involved presenting the results of Experiment #1 (and too an extent, Experiment #2) to the instructor. Below is some of the feedback attained.

It is important to save the December Fair materials. Definitely have them online.

- Reformulate the introduction and stuff
- Maybe omit the southern hemisphere

Abhinav K. Sharma

15 December 2023
12:00 AM - 6:00 AM; 12:30 PM - 2:30 PM

Project Notes Articles #14-20 were all completed, in addition to the three patents. Other aspects of the Project Notes document, including the literature search parameters, knowledge gaps and tags. Much information regarding the use of neural network models, climate time series models, as well as deeper-level parametric models has been attained from reading these studies.

Abhinav K. Sharma

20 December 2023
8:00 AM - 1:00 PM

Notes on Article #21 were taken. This article has specifically formed a major part of the project thought process, as it takes a look at the network that exists among causal relationships of phytoplankton dynamics, identifying how different ecosystems influence which parameters end up being the most influential in phytoplankton population characteristics. It specifically uses chlorophyll as a proxy for both biomass. Moreover, given chlorophyll's role in metabolic pathways, it also acts as an indicator for primary production. Therefore, future data acquisition processes ought to use chlorophyll as a proxy in cases where direct phytoplankton dynamics cannot be observed.

Abhinav K. Sharma

4 January 2024
10:00 AM - 11:00 AM

Time provided in class was used to begin working on the STEM Thesis Introduction.

Abhinav K. Sharma

6 January 2024
2:30 PM - 8:30 PM

Experiment #2 (detailed below) was commenced. This involved carrying out a preliminary analysis of inter-parameter relationships using the same dataset provided from the EPA used for the December Fair. Work on this day involved the creation, troubleshooting, and partial execution of the Python script used for this end.

Abhinav K. Sharma

8 January 2024
4:00 AM - 7:30 AM; 10:30 AM - 11:15 AM; 6:45 PM - 7:45 PM

Work on the STEM Thesis introduction was continued. The execution of the python script for Experiment #2's inter-parameter analysis was fully completed, and statistical results were gathered. Below are some of the observations that were made from the data. The full report is detailed below.

Qualitative Trends Among Inter-Parameter Relationships

- Convergent Homoscedasticity With Sample Size Relationships
    - Est. mean preference among various groups? + Mimics normal curve
- Negative Linear Trends Better Expressed by Exp Decay, in Some cases upward concavity
- Positive Linear Trends Better Expressed by Log Curves, in Some cases upward/downward concavity, one/two cases of exponential growth
- Though considerable amount wouldn't be better off re-expressed, especially for pH
- Temperature Especially fit for log/exponential curve, so does Secchi Disk
- Remember X/Y flip-flop: log→exp; polynomial/concave → radical/conic

Looking at the high R^2 scorers (closer to top of list = higher r^2)

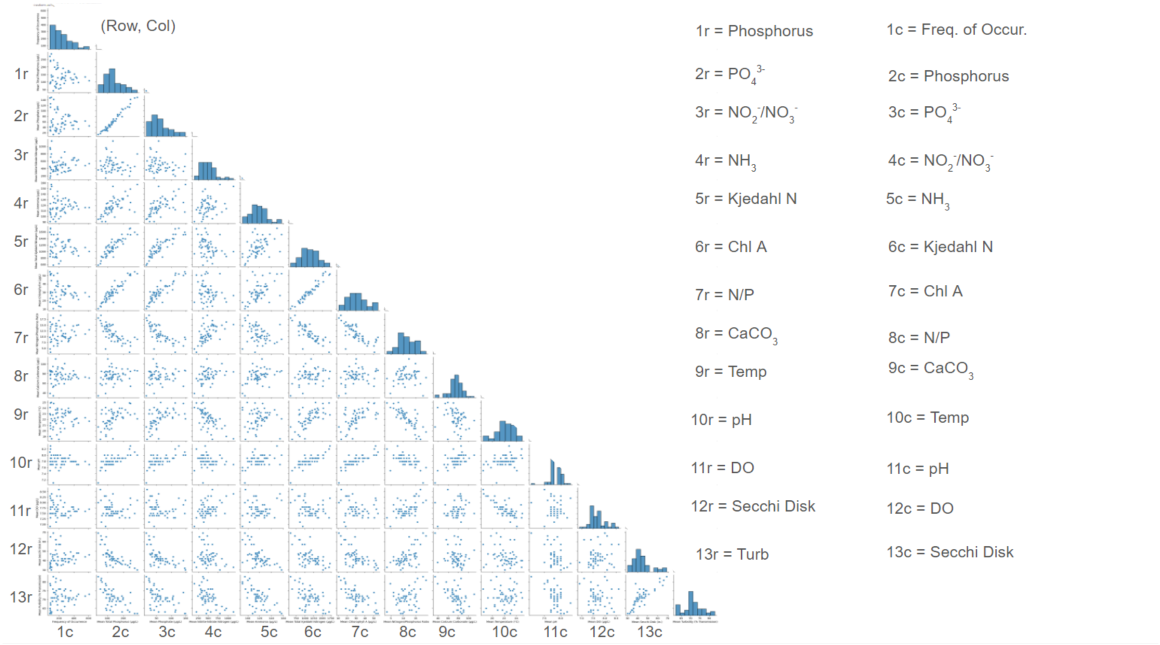
R^2 ≥ 0.7

- Phosphorus vs. Phosphate (intuitive)
- Kjeldahl N vs. Chl A
- Phosphorus vs. Chl A
- Phosphorus vs. Kjeldahl N
- Phosphate vs. Chl A
- Phosphate vs. Kjeldahl N
- Secchi Disk vs. Turb (intuitive)

R^2 ≥ 0.5

- Kjeldahl N vs. pH
- Kjeldahl N vs. N/P Ratio
- N/P Ratio vs. Temp
- Chl A vs. pH
- Temp vs. DO

- Chl A vs. N/P Ratio
- Phosphate vs. N/P Ratio
- Phosphorus vs. N/P Ratio
- Phosphorus vs. pH
- P vs. Secchi Disk (really better with exp decay)



Abhinav K. Sharma 8 January 2024 7:30 PM

A potential diagram to include in Experiment #2 findings. This is a pairplot generated using the Seaborn package.

Abhinav K. Sharma

10 January 2024
7:00 - 7:45 AM; 12:45 - 1:45 PM; 5:00 PM - 6:30 PM

The report on experiment #2 was commenced. In addition, work on the STEM Thesis Introduction was done. Some level of searching for data was achieved, though results have not been satisfactory for project ends. More applicable sources, when found, are to be pasted in a later entry.

Abhinav K. Sharma

11 January 2024
1:30 PM - 2:45 PM

During this work session, the STEM Thesis Introduction was successfully completed.

Abhinav K. Sharma

12 January 2024
4:15 AM - 6:45 AM
A wide variety of potentially applicable data sources were attained and are listed below. Each data source has a status on whether or not it was downloaded.

Data from the NOAA
https://www.ncei.noaa.gov/products/world-ocean-database
Downloaded?: Yes

Data from MassDEP
https://www.mass.gov/guides/water-quality-monitoring-program-data#-data-files-
Downloaded?: No

Data from the USDA
https://www.usda.gov/content/usda-open-data-catalog
Downloaded?: No

Sources and Data from the EPA:
https://www.epa.gov/caddis
Provides a broad overview of the scientific process scientists use for testing water quality. It provides in-depth information on stressors and statistical tools, and also provides examples and case studies.
Downloaded?: *Difficult to do that…*

https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys
This link contains specific csv files collected as part of the National Aquatic Resource Surveys (NARS) on phytoplankton dynamics, as well as a myriad of environmental parameters.
Downloaded?:

Phytoplankton water quality relationships in U.S. lakes: The common phytoplankton genera from eastern and southeastern lakes (note that this source is dated, being from the 1970s, so tread with caution)
Contains the ideal levels for numerous environmental variables for many genera of phytoplankton found throughout the northeastern and southeastern US.

Literature Review on Nutrient-Related Rates, Constants, and Kinetics Formulations in Surface Water Quality Modeling
Provides a series of mathematical models and equations associated with environmental parameters that impact water quality.

https://www.epa.gov/nutrientpollution/nutrient-data
Provides data on pollutants, specifically nitrogen and phosphorus.

Kaggle Datasets:

https://www.kaggle.com/datasets/sohier/calcofi (1)
This dataset contains information about not just phytoplankton and stressors; in fact, the main focus is on fish. Ideally, planktonic data can be extracted, in conjunction with data on fish.
https://www.kaggle.com/code/ankitachoudhury01/water-potability-notebook-seaborn (2)
Although not likely to be used directly for the project (dataset is more about water quality), this is a great tool in that it would act as an example for the necessary Python syntax and statistical tools.
https://www.kaggle.com/datasets/brsdincer/ocean-data-climate-change-nasa/data (3) and
https://www.kaggle.com/datasets/joebeachcapital/global-earth-temperatures (4) provide
General oceanographic data.
Downloaded?: (1) Yes, (2), (3), (4)

Note: No data files were found for most of the links listed below. While it is possible to attain data via requests, the data in most cases below is far too niche for the holistic scale desired for this project.
Potential Data from Article #12 (The Article Utilizing HLM in Wuhan, China)
(Kumar, 2018)
https://www.sciencedirect.com/science/article/pii/S0078323421000981; Suspended Solids
Downloaded? No; no data files found + data too niche for project use

https://www.sciencedirect.com/science/article/pii/S0278434317306477; Micronutrients
Downloaded? No; no data files found + data too niche for project use

(Li et al., 2020)
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231357; Holistic Analysis
Downloaded? No; Data too niche for use in project + data too niche for project use

(Wang et al., 2020)
https://doi.org/10.1016/j.marpolbul.2020.111010; Holistic Analysis + Turbidity, Ammonia Nitrogen, Nitrate Nitrogen
Downloaded? No; no data files found + data too niche for project use

(Jakobsen et al., 2015)
https://www.sciencedirect.com/science/article/pii/S0022098115300083; Light, pH, salinity
Downloaded? No; no data files found + data too niche for project use

(Ke et al., 2019)
https://www.sciencedirect.com/science/article/pii/S1872203217301828; Holistic Analysis + Suspended Solids
Downloaded? No; no data files found + data too niche for project use

(Wu et al., 2019)
https://www.sciencedirect.com/science/article/pii/S1470160X19306685; Transparency and Flow Rate
Downloaded? No; no data files found + data too niche for project use

(Jiang et al., 2014)

https://www.sciencedirect.com/science/article/pii/S1470160X14000089?via%3Dihub; Water Temperature
Downloaded? No; no data files found + data too niche for project use

(Cao et al., 2016)
https://www.sciencedirect.com/science/article/pii/S1470160X15005117?via%3Dihub; Chemical Oxygen Demand, Total Nitrogen
Downloaded? No; although some data is available, it is too niche for project use

(Silva et al., 2014)
https://www.sciencedirect.com/science/article/pii/S0075951114000280?via%3Dihub; Broader ecological scale (focused on hydroelectric settings)
Downloaded? No; although some data is available, it is too niche for project use

Abhinav K. Sharma

13 January 2024
8:15 AM - 1:00 PM; 1:30 PM - 9:45 PM

After diligent research, the final decision was made on what dataset to use. The 2018 World Ocean Database's Ocean Station Dataset provided by the NOAA was attained using Amazon Web Services S3 explorer. NetCDF files were downloaded. This was a time-consuming process that took about 8 hours over the course of two days. Indeed, 25.8 GB of data in total was downloaded. Moreover, the Panoply software downloaded and used for performing metadata analysis so a proper image of the data could be attained before extracting raw values.

Abhinav K. Sharma

14 January 2024
8:45 AM - 12:30 PM; 1:00 PM - 7:45 PM
The downloading of data was completed. The files extraction process from NetCDF files to .csv files was commenced. The plan for this process was established through creating a Python script responsible for appending each year's dataset to a list. Then, for each year and each parameter, the data were to be extracted. To help get a basic idea of parametric data, notes on some basic attributes, including sample size, time frame, and units of measurement were taken below:

Parameters from NOAA's OSD Data:

- Temp (℃) **1772-2017** (n = 2,845,911)
- Salinity (standard salinity units) **1873-2017** (n = 2,408,713)
- DO (μmol/kg) **1878-2017** (n = 913,215)
- PO_4^{3-} (μmol/kg) **1922-2017** (n = 597,499)
- SiO_4^{4-} (μmol/kg) **1921-2017** (n = 461,801)
- NO_3^-/NO_2^- (μmol/kg) **1925-2017** (n = 372,557)
- pH (H^+ ions/unitless) **1910-2017** (n = 265,898)

- Chlorophyll (µg/L) **1933-2017** (n = 220,059)
- Alkalinity ($CaCO_3$ - milli-equivalent/liter) **1921-2017** (n = 71,932)
- pCO_2 (µatm)**1967-2014** (n = 3,086)
- DIC (mmol/L)**1958-2017** (n = 21,588)
- $^{3}_{1}$H (Tritium - Tritium Unit) **1984-2015** (n = 1,876)
- He (nmol/kg) **1984-2013** (n = 1,979)
- △^{3}He (%)**1985-2013** (n = 1,113)
- △^{14}C (‰ deviation) **1987-2014** (n = 1,726)
- △^{13}C (‰ deviation)**1991-2014** (n = 1,800)
- ~~Argon~~ (nmol/kg)-**1993** (omitted due to negligible timeframe, plus n = 73 only)
- Neon (nmol/kg)**1987-2013** (n = 1,381)
- Chlorofluorocarbon-11 (pmol/kg) **1987-2013** (n = 16,530)
- Chlorofluorocarbon-12 (pmol/kg) **1987-2013** (n = 16,617)
- Chlorofluorocarbon-13 (pmol/kg) **1987-2013** (n = 6,706)
- △^{18}O (‰ deviation) **1987-2013** (n = 1,186)
- Pressure (decibars) **1987-2013** (n = 207,107)
- Tax and biomass of Phytoplankton, Zooplankton and Ichthyoplankton **1900-2015** (n = 245,059)

Abhinav K. Sharma

15 January 2024
8:15 AM - 1:30 PM; 2:00 PM - 7:45 PM

Once the basic setup for the Python code was done, the next step was to execute it. This was a long and arduous process. The first parameters that were to undergo this process were temperature and chlorophyll. Two key realizations were made as metadata analysis was performed: neither direct geospatial nor phytoplankton data could be attained. Therefore, an approach was taken wherein total oceanic chlorophyll was used as an indicator for phytoplankton biomass and primary production (based off research from Article #21). Moreover, each year's average, standard deviation, and sample size for each parameter was taken and collated. This basic process was foundational to the overall procedure and data analysis for this project, and was to be carried out.

Abhinav K. Sharma

16 January 2024
4:00 AM - 7:30 AM; 1:30 PM - 4:30 PM

Work on this day was dedicated to the 7th STEM Update meeting. This involved the presentation of the preliminary model establishment (i.e., the conclusions reached in the previous entry).

Abhinav K. Sharma

17 January 2024
2:00 AM - 5:00 AM; 3:00 PM - 4:30 PM

From the procedural foundation established from the past days, the process of applying this to all other parameters was begun. The python code described above was created using Google Colaboratory. Uploading netCDF files was an exhaustive process that required constant troubleshooting with internet connection as well as IDE and local device stability, as the size of the data was of a magnitude that was very hard for the computer to handle.

Abhinav K. Sharma

24 January 2024
2:30 AM - 6:15 AM; 7:00 AM - 9:45 AM; 6:00 PM - 9:00 PM
This process of data acquisition described above was continued.

Abhinav K. Sharma

25 January 2024
2:00 AM - 7:00 AM; 12:30 PM - 2:45 PM; 5:00 PM - 9:00 PM
This process of data acquisition described above was continued.

Abhinav K. Sharma

26 January 2024
3:30 AM - 7:00 AM; 4:00 PM - 11:00 PM
This process of data acquisition described above was continued.

Abhinav K. Sharma

27 January 2024
9:45 AM - 12:15 PM; 1:00 PM - 10:00 PM
This process of data acquisition described above was continued.

Abhinav K. Sharma

28 January 2024
12:00 AM - 1:00 AM; 8:00 AM - 1:30 PM; 3:00 PM - 6:15 PM
This process of data acquisition described above was continued.

Abhinav K. Sharma

29 January 2024
5:00 PM - 9:00 PM
This process of data acquisition described above was completed. In the meantime, a process for performing statistical analysis had been formed. This involved sinusoidal regression for the time series models of parametric values, linear regression models of parameters with the chosen indicator, and PCA

and covariance matrix to identify driving parameters and inter-parameter relationships. These analyses had been performed in tandem with the above entries.

Abhinav K. Sharma

30 January 2024
7:00 AM - 7:45 AM; 6:00 - 8:00 PM
Work sessions on this day were dedicated to beginning a draft of an entire research paper based off the above described procedure. Experiment #3 and onward below cover the scope of the paper. This was done for the Junior Symposium of Humanities and Science (JSHS) competition.

Abhinav K. Sharma

31 January 2024
3:00 AM - 6:45 AM; 4:00 PM - 11:59 PM
The above described research paper was completed. This included a full and complete abstract, introduction, methodology, results, and conclusions section.

Abhinav K. Sharma

1 February 2024
7:00 - 7:45 AM

Using the draft of the completed paper for JSHS, various STEM class assignments related to drafting a methodology, results, and conclusions section were completed.

Abhinav K. Sharma

5 February 2024
1:45 - 2:45 PM

This class time was used to finalize the abstract of this paper, and submit it to MSEF.

Abhinav K. Sharma

6 February 2024
4:00 PM - 7:00 PM

Work on this day was dedicated to preparing for the Final STEM Update meeting. This involved the translation of the draft of my paper into a slideshow format. The work done here has served as a preliminary basis whereby materials for the February Fair can be collated.

Abhinav K. Sharma

7 February 2024
2:00 AM - 7:00 AM; 11:45 AM - 12:15 PM

Update meeting preparations were continued. The final STEM update meeting was then attended. While overall presentation content was sound, a notable problem was the excessive length of the presentation. More effort needs to be placed in concisely presenting material.

Abhinav K. Sharma

10 February 2024
10:00 AM - 12:00 PM; 12:30 PM - 8:00 PM

The poster for the February Fair was developed. In tandem with this, all STEM Website requirements were complete. This involved the completion of the STEM Logbook, as well as all other products necessary for this project. Two emails were written to try and advance this project professionally. These can be found in the above professional communications section.

Abhinav K. Sharma

11 February 2024
7:30 AM - 12:00 PM; 12:30 PM - 11:00 PM

The poster for the February Fair was developed. In tandem with this, all STEM Website requirements were complete. This involved the completion of the STEM Logbook, as well as all other products. In other words, all requisites for completing the STEM project were satisfied. Definitely proud of all that I have completed.

Abhinav K. Sharma

12-14 February 2024

Work on these days were dedicated to practicing presenting for the February Fair.

Abhinav K. Sharma

15 February 2024

The February Fair was attended, and the presentation of the project was carried out. The entire project process has been enjoyable, albeit frustrating, overwhelming, and stressful at times. I feel as though the work that I have undergone has allowed me to grow my research, problem-solving, critical thinking and analysis skills, as well as my abilities with professional correspondence and public speaking. I hope to continue doing projects like these in the future, and I am excited to see where this project takes me!

Experiments:

Experiment 1:
Preliminary Parameter Analysis
7 December 2023
Abhinav K. Sharma

Introduction:

The first empirical procedure performed for this project involved performing a series of One-Way ANOVA tests and Post-HOC Tukey tests on a dataset attained from the EPA. The aim was to determine, among the fifty-seven genera listed in the sample, whether there were any differences in mean parameter values. For example, one of the fourteen parameters tested was mean water temperature for each genera. The above-mentioned statistical tests were then used to test if mean temperature values found in each genera significantly varied, and where (that is, which specific genera) that variation could be found. Findings indicate that mean factor values vary significantly among genera, while others do not. These findings may be useful for making decisions in initial model development.

Methods/Materials

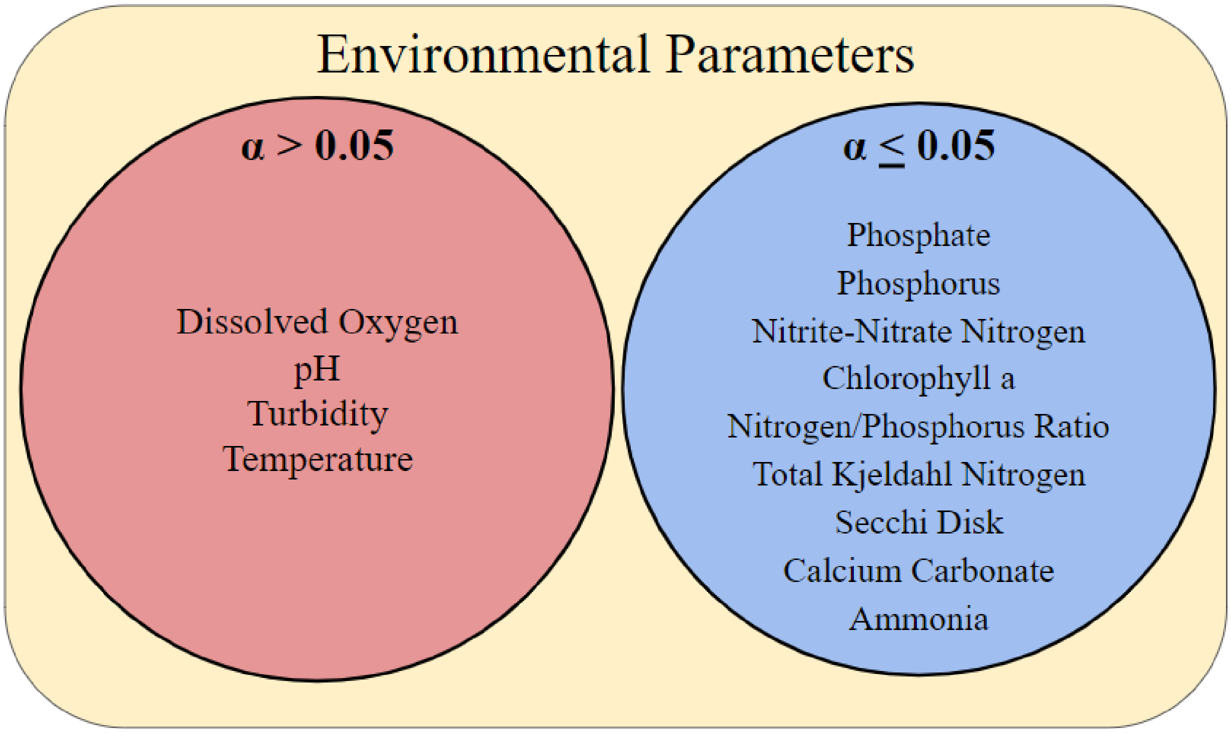
A dataset from the EPA (Hern et al., 1979) was attained. The data used was taken from Table 3 of the original paper. Using an online convert of image to tabular data, these data were collated into a Google Sheets file (linked below). The original dataset was reorganized such that each genus was listed in alphabetical order, followed by their frequency of observation and mean parameter values. Sample standard error was attained by dividing each mean parameter value for each genus by the square root of the frequency. Then, the data was further reorganized into multiple sheets for each parameter. Using this online software, one-way ANOVA, along with post-hoc Tukey tests were performed using summary statistics (sample size, mean, standard error). This software not only produced individual pair results for the post-hoc tests, but it also provided a series of 95% confidence intervals for the spread of the environmental factor for each genus of phytoplankton, as well as ANOVA tables. Each parameter in the spreadsheet file linked below

Observations and Experimental Data:

A more detailed view of the data is offered by the spreadsheet below. The following figures offer a summary of the overall trends.

Figure 2

Assortment of Insignificant and Significant Parameters



Note. The yellow box is used to denote the set of all environmental parameters. Within this sampling space, the red and blue circles depict the insignificant and significant parameters, respectively (α = 0.05).

Table 1

One-Way ANOVA and Post-Hoc Tukey Test Results for Parameters

Parameter	Significance Level For One-Way ANOVA	Proportion of Significantly Different Pairs (α = 0.05) From Post-HOC Tukey Tests
Phosphate (µg/L)	0	0.5432
Phosphorus (µg/L)	2.11E-220	0.4467
Nitrite-Nitrate Nitrogen (µg/L)	2.04E-192	0.3690
Chlorophyll a (µg/L)	6.53E-163	0.3766
Nitrogen/ Phosphorus Ratio	3.13E-128	0.3045
Total Kjeldahl Nitrogen (µg/L)	1.05E-64	0.1823
Secchi Disk (in.)	9.25E-42	0.1397
Calcium Carbonate (µg/L)	7.72E-31	0.0877
Ammonia (µg/L)	2.59E-06	0.0025
Temperature (℃)	8.84E-01	0.0000
Turbidity (% Trans.)	9.16E-01	0.0000

DO (µg/L)	1.00E+00	0.0000
pH	1.00E+00	0.0000

Note. Parameters are listed top to bottom in order of decreasing significance level (greater p-value). A color gradient from blue to red accompanies this progression. Since fifty-seven genera were tested, there were 1,596 unique pairs to compare in the Post-HOC Tukey tests. The proportions in the rightmost column are thus out of 1,596. To aid in clarity, a visual was provided at the left of each parameter.

Calculations and Data Analysis:

Overall Trends:

- There are more factors that have significant inter-genera variation than those that do not
    - Parameters with insignificant results from the ANOVA tests can be considered homogenous among phytoplankton groups, whereas those with significant results from the ANOVA tests were considered heterogeneous.
- There is a negative correlation between significance level and proportion of significant pairs from Post-HOC Tukey Tests. That is, the number of significant pairs from the Post-HOC Tukey tests decreased the greater the p-value, and hence, the more homogenous the parameter distribution was to be found.
- Most significance level values are either near 0 or near 1
- There is substantial divergence from the inter-genera mean for significantly different parameters, whereas there is little divergence in non-significant ones as can be seen from the 95% confidence intervals. These dynamics follow a sort of gradient among parameters as significance level decreases.
- ANOVA tables support respective results.

Anomalies

- There is one notable departures from negative correlation between significance level and proportion of significant pairs from Post-HOC Tukey Tests
    - Nitrite-Nitrate Nitrogen, which has 20 CI's containing inter-genera mean vs. Chlorophyll a, which has 18 CI's containing inter-genera mean
- Ammonia (Table 1)
    - 9 CI's w/ small deviations

Concluding Remarks

This data and analysis have a few limitations given that the timeframe of when the data were collected (1970s), and that the scope was small, only focusing on fifty-seven genera in the Eastern United States. However, this data serves as a solid preliminary investigation of what parameters to use for initial model development and validation. To these ends, parameters found to be homogenous are likely to be better candidates than those that are heterogenous. This is because having less varied data initially will make model construction and validation easier. Moreover, homogenous factors, when changed, will likely impact a wider range of phytoplankton in a more consistent way than factors that vary. Hence, preliminary predictive modeling will also be easier using such parameters.

The spreadsheet file linked below contains the specific experimental data.
2023_STEMProject_Exp1_Data_Sharma_v1

Experiment 2:

Preliminary Analysis of Inter-Parameter Relationships
8 January 2024

Abhinav K. Sharma

Introduction:

This empirical analysis was performed with the goal of providing a preliminary look at the relationship between different parameters. Understanding their interactions is crucial, as all of these parameters play a major role in influencing phytoplankton dynamics. As these values change due to global warming, in turn modifying phytoplankton characteristics, knowing the possible underlying relationship between parameters can provide information on how their levels might change in tandem, and in turn, how that may influence the net impact on phytoplankton populations. Therefore, using the same dataset from the previous experiment, the relationship between every combination of parameters was assessed by applying linear regression. Using the pairplot function from the seaborn package was used as a frame of comparison to verify the accuracy of the individual linear models being produced.

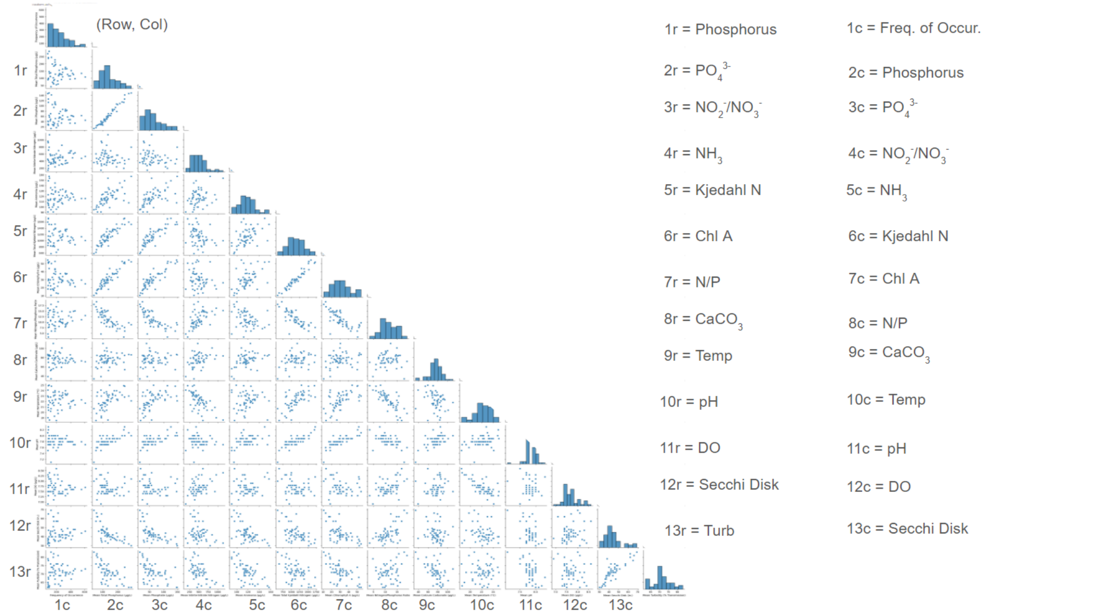
Methods/Materials:

Using the same EPA dataset from experiment 1, a series of linear regression models of inter-parameter relationships was developed. This was achieved using the python code linked below. It is important to note that the correlative relationships between parameters indicate the relationship among the preferences of the different genera of phytoplankton. That is, a negative correlation between two parameters in this context would suggest the following: given a preference for a higher level of one parameter, it would be predicted that there would be a preference for a lower level of the other parameter. Therefore, there is no direct causation between the parameters themselves. Rather, the correlations among the direction of the general preferences are being observed.

Observations and Experimental Data:

Figure 1

Pairplot of Inter-parameter Relationships



Note. The key on the right denotes which parameter corresponds to which row or column along the pairplot.

Table 1

Structure of Parametric Combinations

	Frequency of Occurrence	Mean Total Phosphorus (µg/L)	Mean Phosphate (µg/L)	Mean Nitrite-Nitrate Nitrogen (µg/L)	Mean Ammonia (µg/L)	Mean Total Kjeldahl Nitrogen (µg/L)	Mean Chlorophyll A (µg/L)	Mean Nitrogen/Phosphorus Ratio	Mean Calcium Carbonate (µg/L)	Mean Temperature (°C)	Mean pH	Mean DO (µg/L)	Mean Secchi Disk (in.)	Mean Turbidity (% Transmission)
Frequency of Occurrence	x	1	2	3	4	5	6	7	8	9	10	11	12	13
Mean Total Phosphorus (µg/L)	1	x	14	15	16	17	18	19	20	21	22	23	24	25
Mean Phosphate (µg/L)	2	14	x	26	27	28	29	30	31	32	33	34	35	36
Mean Nitrite-Nitrate Nitrogen (µg/L)	3	15	26	x	37	38	39	40	41	42	43	44	45	46
Mean Ammonia (µg/L)	4	16	27	37	x	47	48	49	50	51	52	53	54	55
Mean Total Kjeldahl Nitrogen (µg/L)	5	17	28	38	47	x	56	57	58	59	60	61	62	63
Mean Chlorophyll A (µg/L)	6	18	29	39	48	56	x	64	65	66	67	68	69	70
Mean Nitrogen/Phosphorus Ratio	7	19	30	40	49	57	64	x	71	72	73	74	75	76
Mean Calcium Carbonate (µg/L)	8	20	31	41	50	58	65	71	x	77	78	79	80	81
Mean Temperature (°C)	9	21	32	42	51	59	66	72	77	x	82	83	84	85
Mean pH	10	22	33	43	52	60	67	73	78	82	x	86	87	88
Mean DO (µg/L)	11	23	34	44	53	61	68	74	79	83	86	x	89	90
Mean Secchi Disk (in.)	12	24	35	45	54	62	69	75	80	84	87	89	x	91
Mean Turbidity (% Transmission)	13	25	36	46	55	63	70	76	81	85	88	90	91	x

Note. This table illustrates the numerical assignment of parametric combinations to all possible combinations. Blue boxes highlight inverses of parametric combinations that were omitted, as those functions would not present any information outside of an inverse function, which, for the purposes of this experiment, would be redundant.

Table 2

Summary Statistics of Parametric Combinations

Combination	p-value	R^2	Notes/Observations

1	0.3406	0.02	Convergent Heteroscedasticity (slightly neg)
2	0.3100	0.02	Convergent Heteroscedasticity (slightly neg)
3	0.7312	0.00	Convergent Heteroscedasticity
4	0.8411	0.00	Convergent Heteroscedasticity
5	0.3849	0.01	Convergent Heteroscedasticity (notably more neg)
6	0.2739	0.02	Convergent Heteroscedasticity (notably more neg)
7	0.3458	0.02	Convergent Heteroscedasticity (notably more pos)
8	0.9971	0.00	Convergent Heteroscedasticity
9	0.8550	0.00	Convergent Heteroscedasticity
10	0.4365	0.01	Convergent Heteroscedasticity (slightly neg)
11	0.3012	0.02	Convergent Heteroscedasticity (notably more neg)
12	0.8292	0.00	Convergent Heteroscedasticity
13	0.5654	0.01	Convergent Heteroscedasticity (slightly pos)
14	0.0000	0.97	Strong Positive Linear
15	0.4863	0.01	Slightly Neg, Large Spread, Bimodal outlier on either side
16	0.0000	0.41	Strong Positive, Divergent Heteroscedasticity
17	0.0000	0.80	Fairly Strong Positive, Sight Log curve
18	0.0000	0.86	Strong Positive Linear
19	0.0000	0.52	Fairly Strong Negative
20	0.1127	0.05	Weak Pos, Large Spread, Slight Conv HetSced
21	0.0000	0.26	Weak POs, Large Spread, Log Curve
22	0.0000	0.50	Weak Pos, Moderate Scatter
23	0.4082	0.01	Weak Neg, Upward Concave
24	0.0000	0.50	Moderate Neg, Exp Decay
25	0.0000	0.40	Weak Neg, Maybe Exp Decay
26	0.1787	0.03	Slightly Neg, Large Spread, Bimodal outlier on either side
27	0.0000	0.36	Weak Pos, Large Spread, Slight Conv HetSced, Log Curve
28	0.0000	0.76	Moderate Pos, Maybe Log Curve
29	0.0000	0.80	Moderate Pos
30	0.0000	0.54	Weak Neg, Exp Decay
31	0.3814	0.01	Slight Pos, High Spread, Slight Conv HetSced
32	0.0000	0.30	Weak Pos, Log Curve
33	0.0000	0.44	Weak Pos
34	0.3768	0.01	Slight Pos, High Spread, Upward Concave

35	0.0000	0.40	Weak neg, Exp Decay
36	0.0001	0.26	Weak neg, exp Decay
37	0.0116	0.11	Weak pos, slight div HetSced
38	0.0160	0.10	Weak Neg, Conv HetSced, Exp Decay
39	0.1216	0.04	Weak Neg, Conv HetSced, Exp Decay
40	0.0000	0.31	Weak Pos, Conv HetSced, Maybe Log Curve
41	0.0040	0.14	Weak Pos, Conv HetSced, Maybe Log Curve
42	0.0000	0.45	Weak Neg, Conv HetSced, Maybe Exp Decay
43	0.3636	0.02	Slight Neg, Conv HetSced
44	0.0002	0.23	Weak pos, maybe log curve
45	0.0336	0.08	Weak neg, Conv HetSced, exp curve
46	0.0001	0.26	Weak neg, Conv HetSced, exp curve
47	0.0000	0.26	Weak Pos, Log Curve
48	0.0000	0.27	Weak pos, Div HetSced, Maybe Log curve
49	0.0130	0.11	Weak Neg, Maybe Exp Decay
50	0.0017	0.17	Weak Pos, Log Curve
51	0.5579	0.01	Weak Pos, Log Curve
52	0.0002	0.22	Weak Pos
53	0.4087	0.01	Weak Pos, Exp Decay or Maybe Upward Concave
54	0.0000	0.37	Weak neg, exp Decay
55	0.0000	0.34	Weak Neg, Maybe Exp Decay
56	0.0000	0.93	Strong Positive Linear
57	0.0000	0.65	Moderate Neg
58	0.0431	0.07	Weak Pos, Maybe downward concave or log curve
59	0.0000	0.43	Weak Pos, Log Curve
60	0.0000	0.67	Moderate Pos
61	0.0776	0.06	Weak neg, Conv-->Div HetSced, Maybe Exp decay or upward concave
62	0.0000	0.36	Weak Neg, Exp Decay
63	0.0000	0.27	Weak neg
64	0.0000	0.55	Moderate neg, maybe exp decay
65	0.0425	0.07	Weak pos, Log curve, maybe downward concave
66	0.0000	0.34	Weak pos, Log curve,
67	0.0000	0.62	Moderate Pos
68	0.2654	0.02	Weak neg, maybe Exp Decay or upward concave

69	0.0000	0.39	Weak Neg, Exp Decay
70	0.0000	0.36	Weak Neg, Exp Decay
71	0.3869	0.01	Weak pos, Log curve, maybe downward concave
72	0.0000	0.63	Mod Neg, Div HetSced, maybe exp decay
73	0.0000	0.32	Weak Neg
74	0.0002	0.23	Weak Pos
75	0.0000	0.34	Weak Pos, exp growth
76	0.0398	0.07	Weak Pos, upward concave
77	0.0308	0.08	Weak neg
78	0.0000	0.31	Weak Pos
79	0.1391	0.04	Weak pos, central cluster
80	0.0757	0.06	Weak neg
81	0.0000	0.27	Weak Neg, Slight Div HetSced
82	0.0030	0.15	Weak pos
83	0.0000	0.56	Weak Neg, exp Decay
84	0.0350	0.08	Weak Neg, exp Decay
85	0.7805	0.00	Weak pos, Upward concave
86	0.7111	0.00	Weak neg
87	0.0001	0.25	Weak neg
88	0.0000	0.26	Weak neg
89	0.2333	0.03	Weak pos, maybe exp growth
90	0.4862	0.01	Weak neg, maybe upward concave
91	0.0000	0.71	Moderate pos

Note. This table illustrates each parametric combination with its R^2, p values, and the qualitative relationships between the parameters. The actual graphs may be found in the spreadsheet linked below.

Calculations and Data Analysis:
Qualitative Trends Among Inter-Parameter Relationships

- Convergent Homoscedasticity With Sample Size Relationships
    - Est. mean preference among various groups? + Mimics normal curve
- Negative Linear Trends Better Expressed by Exp Decay, in Some cases upward concavity
- Positive Linear Trends Better Expressed by Log Curves, in Some cases upward/downward concavity, one/two cases of exponential growth
- Though considerable amount wouldn’t be better off re-expressed, especially for pH
- Temperature Especially fit for log/exponential curve, so does Secchi Disk
- Remember X/Y flip-flop: log→exp; polynomial/concave → radical/conic

Looking at the high R^2 scorers (closer to top of list = higher r^2)

R^2 ≥ 0.7

- Phosphorus vs. Phosphate (intuitive)
- Kjeldahl N vs. Chl A
- Phosphorus vs. Chl A
- Phosphorus vs. Kjeldahl N
- Phosphate vs. Chl A
- Phosphate vs. Kjeldahl N
- Secchi Disk vs. Turb (intuitive)

R^2 ≥ 0.5

- Kjeldahl N vs. pH
- Kjeldahl N vs. N/P Ratio
- N/P Ratio vs. Temp
- Chl A vs. pH
- Temp vs. DO
- Chl A vs. N/P Ratio
- Phosphate vs. N/P Ratio
- Phosphorus vs. N/P Ratio
- Phosphorus vs. pH
- P vs. Secchi Disk (really better with exp decay)

Conclusions that can be drawn from this:

Many of the correlations identified above are intuitive. For example, phosphorus and phosphate both have the same base element of phosphorus. It would follow, therefore, that preference in increased phosphate levels corresponds to preferences of increased phosphorus level. It appears that there are strong correlations in the level of preferences among chlorophyll a and various nutrients. Indeed, in being parameters related to nutrient cycling and metabolism, strong connections would follow. However, less obvious relationships seem to appear here: for example, with pH, temperature, DO, as well as with P vs. Secchi Disk, parameters divorced of common biochemical processes as intimate as metabolism and biogeochemical nutrient cycling.

Concluding Remarks:

Overall, this data has provided a wide variety of notable results in terms of the sheer magnitude of significant inter-parametric relationships observed. While some correlative relationships are fairly intuitive to understand, such as those between nutrients that derive from the same molecule, others are not so obvious, and warrant further study. Notably, the relationships between pH and nutrients warrant further study, as their connection appears to be less intuitive from a chemical perspective than that e.g., chlorophyll, nitrogen, and phosphorus, all of which are nutrients and chemicals involved in interconnected metabolic processes. Meanwhile, pH is an outside actor. Perhaps investigating these dynamics would result in ramifications with regards to metabolism as a whole. As a whole, while this experiment may not be directly connected to the project goal, it has nonetheless offered a useful initial look into the nature of inter-parameter relationships. Studying these relationships further in-depth may reveal insights in areas other than that of the scope of this project.

2324_STEMProject_Exp1_Data_Sharma_v1

2324_STEMProject_Exp1_Code_Sharma_v1.py

Comments on Code:

- Read the CSV file imported
- Declare the genera as category
- Establish all parameters as data frame variables
- Create a pairplot using Seaborn package
- Ignore all code below that. That is merely for preliminary testing with linear model formation.

Experiment 3:

Preliminary Testing of Final Computational Apparatus

16 January 2024

Introduction:

This is to be prefaced by the dataset attained. The 2018 World Ocean Database (WOD18) is the most comprehensive datasets on oceanographic data, the most temporally and geographically cosmopolitan of which is the Ocean Station Dataset (OSD). With such a strong empirical basis, using this dataset posed as a great advantage in terms of model applicability and the ramifications of results. Thus, applying the computational apparatus of model validation of the identification, characterization, and analysis of driving parameters, inter-parameter relationships, impact on phytoplankton dynamics, and in turn, climatic and ecological ramifications, and observing the results, would be highly potent. Successful methodology would indicate that this procedure could be continued for more parameters for this dataset, in turn providing major ramifications for oceanic, ecological and climatic conditions at a large scale.

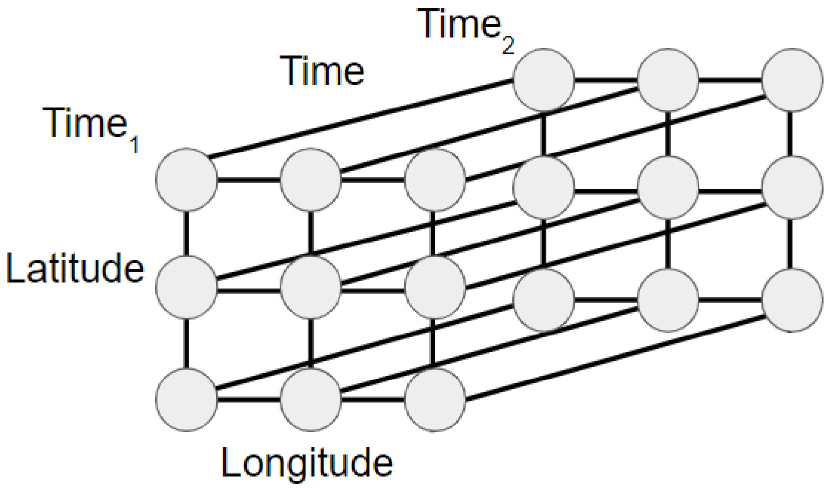
Methods/Materials:

The primary dataset analyzed in this study was the 2018 World Ocean Database (WOD18) provided by the National Oceanic and Atmospheric Administration (NOAA). Both spatially and temporally, this dataset provides a highly cosmopolitan measurement of numerous environmental parameters, including water temperature, micronutrients, pH, salinity, among many others (Boyer et al., 2018). Access to the files of this dataset was attained through the Registry of Open Data provided by Amazon Web Services (AWS). The files used in this dataset were all updated within the AWS S3 explorer system on 17 October 2023 when obtained. Files were organized by year from 1900 to 2023, with pre-1900 data being referred to as 1800 (Amazon Web Services, 2024). The file of each year was systematically downloaded. Regression models were developed using Python code. The main programming interface used was Google Colaboratory, which, when used, was most recently updated on 8 January 2024, supporting Python 3.10.12 (Google Colaboratory, 2024). Across all Python programs developed, the Pandas, Xarray, NumPy, SciPy, SciKitLearn, Seaborn, and MatPlotLib packages were utilized. Additionally, to observe the dimensions of the data files more closely, the Panoply software, provided by the National Aeronautic and Space Administration's (NASA's) Goddard Institute for Space Studies (GISS), was downloaded. The most current version, 5.3.1, was used, released 1 January 2024 (NASA GISS, 2024)

Two key observations were made during preliminary use and analysis of the dataset that led to two key decisions on the analytical procedure performed. The first observation relates to the data structure of the files used. These were NetCDF (.nc) files, which illustrate parametric measurements at specific points of latitude, longitude, and depth across a series of equal temporal increments (Figure 1). In the metadata attained for the .nc files, these spatiotemporal attributes were referred to as "coordinates." However, it was specifically noted within this directory that all dimensions, with the exception of the numerical measurements of parametric data, had their latitude, longitude, depth, and time alongside their measurement. Meaning, outside of the year as provided by the file, the spatiotemporal attributes of the primary data on environmental parameters could not be accessed. Therefore, a more holistic approach to analysis was taken, wherein the average global value of each factor was calculated for each year of data. This helped preserve temporal analysis and offered a viable overview of environmental relationships, though at the expense of omitting the nuances and consequences caused by dissimilar spatial attributes among parametric data.

Figure 1

Illustration of the NetCDF Data Structure

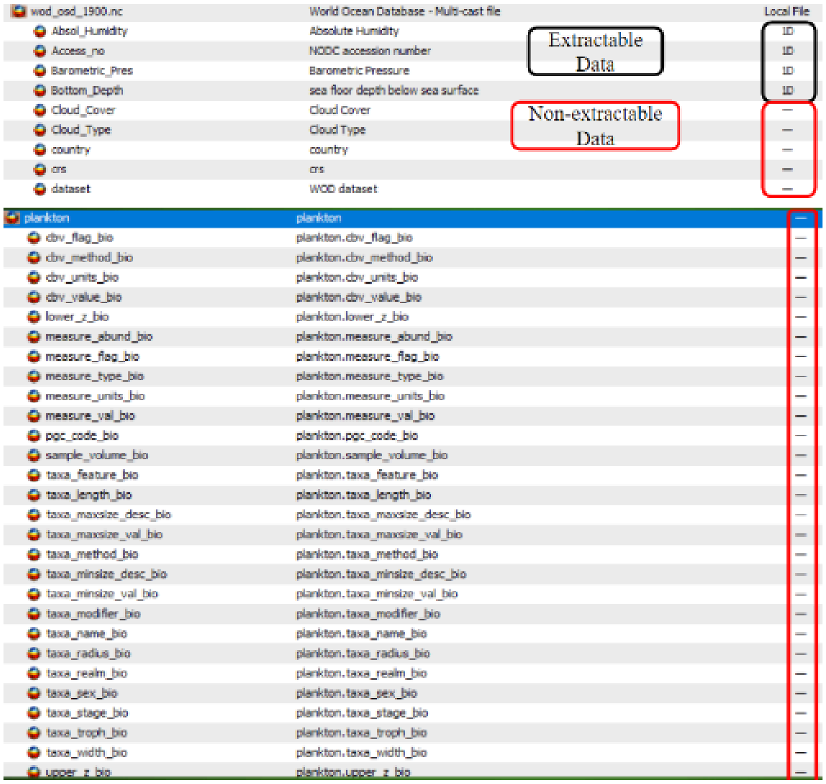


Note. This is a common illustration of the structure of a NetCDF file. The gray circles represent specific points in longitudinal and latitudinal space. Though not illustrated above, each of these coordinates also contains varying levels of depth. At all of these points in space, there exist parametric measurements. These measurements are projected out through a series of time increments, forming a three-dimensional figure.

The second observation made from the metadata was that planktonic data was inaccessible. Within the Panoply software, whereas extractable data was stored within one-dimensional arrays, the values of non-extractable data were unavailable. Figure 2 contains an illustration of this issue. Planktonic data fell under the latter category. Consequently, it was decided to use average global oceanic concentration of total chlorophyll as an indicator of phytoplankton dynamics, namely primary production. This is because chlorophyll is a crucial pigment for carrying out the photosynthetic process, and in turn, all other metabolic processes. As such, a higher concentration of chlorophyll would indicate greater potential for primary production, whereas Though a valuable indicator, it is important to note that it is not a direct measurement of phytoplankton traits. Although data of all available parameters spanning 1900 to 2019 were downloaded, due to the limited temporal range of measurements for total oceanic chlorophyll, this study focused on data from 1954 to 2017. This also reduced the number of factors assessed.

Figure 2

Inaccessibility of Planktonic Data



wod_osd_1900.nc	World Ocean Database - Multi-cast file	Local File
Absol_Humidity	Absolute Humidity	1D
Access_no	NODC accession number	1D
Barometric_Pres	Barometric Pressure	1D
Bottom_Depth	sea floor depth below sea surface	1D
Cloud_Cover	Cloud Cover	—
Cloud_Type	Cloud Type	—
country	country	—
crs	crs	—
dataset	WOD dataset	—
plankton	plankton	—
cbv_flag_bio	plankton.cbv_flag_bio	—
cbv_method_bio	plankton.cbv_method_bio	—
cbv_units_bio	plankton.cbv_units_bio	—
cbv_value_bio	plankton.cbv_value_bio	—
lower_z_bio	plankton.lower_z_bio	—
measure_abund_bio	plankton.measure_abund_bio	—
measure_flag_bio	plankton.measure_flag_bio	—
measure_type_bio	plankton.measure_type_bio	—
measure_units_bio	plankton.measure_units_bio	—
measure_val_bio	plankton.measure_val_bio	—
pgc_code_bio	plankton.pgc_code_bio	—
sample_volume_bio	plankton.sample_volume_bio	—
taxa_feature_bio	plankton.taxa_feature_bio	—
taxa_length_bio	plankton.taxa_length_bio	—
taxa_maxsize_desc_bio	plankton.taxa_maxsize_desc_bio	—
taxa_maxsize_val_bio	plankton.taxa_maxsize_val_bio	—
taxa_method_bio	plankton.taxa_method_bio	—
taxa_minsize_desc_bio	plankton.taxa_minsize_desc_bio	—
taxa_minsize_val_bio	plankton.taxa_minsize_val_bio	—
taxa_modifier_bio	plankton.taxa_modifier_bio	—
taxa_name_bio	plankton.taxa_name_bio	—
taxa_radius_bio	plankton.taxa_radius_bio	—
taxa_realm_bio	plankton.taxa_realm_bio	—
taxa_sex_bio	plankton.taxa_sex_bio	—
taxa_stage_bio	plankton.taxa_stage_bio	—
taxa_troph_bio	plankton.taxa_troph_bio	—
taxa_width_bio	plankton.taxa_width_bio	—
upper_z_bio	plankton.upper_z_bio	—

Note. The extractable data (black) were stored as one-dimensional arrays, whereas the values of inextricable data could not be obtained. All planktonic data fell under the latter category.

Data Extraction and Cleaning

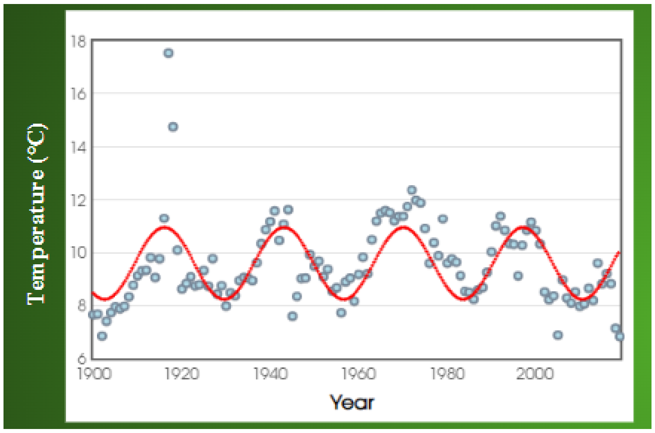
Using Google Colaboratory, a brief program was written to extract parametric data .nc files and save them as Comma Separated Value (.csv) files. Within each .csv file, the average, standard error, and sample size for each year of each parameter was calculated and compiled into a separate spreadsheet file. The main data cleaning involved the removal of non-numerical data within .nc files when converting them to .csv. This was achieved by the Python program written.

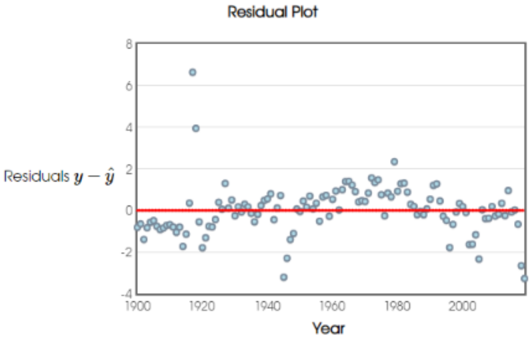
However, for total oceanic chlorophyll and alkalinity, the data processing was more complex. When originally creating a time series for the former, it was noted that model strength was inhibited by abnormally high measurements around the early 2000s. Upon further investigation of the .csv files, it was noted that this was due to the abnormally high amount of outliers. In order to maximize model fitness, for chlorophyll data spanning 1998 to 2008, any and all measurements in excess of 20 µg/L were removed, and new averages, standard deviations, and sample sizes were determined. The next iteration of the time series had a stronger fit as a result. For alkalinity, many years had errors in how the data was recorded, in that the decimal place was improperly positioned. This led to values that were orders of magnitude too high for the dataset, and in turn, skewed summary statistics. As such, any such data was eliminated from the set, with summary statistics adjusted accordingly. Besides the processes described, all parametric data from 1954-2017 was preserved when performing data analysis.

Observations and Experimental Data:

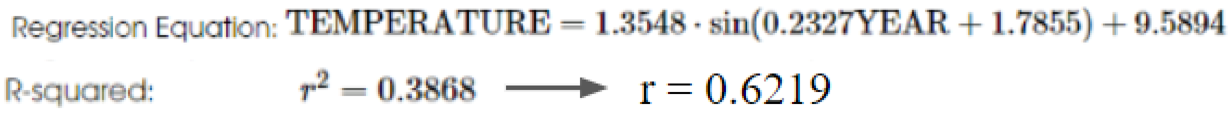
Figure 1

Time Series of Average Global Oceanic Temperature (℃) with Residual Plot





Note. Preliminary sinusoid for average oceanic temperature with residual plot.

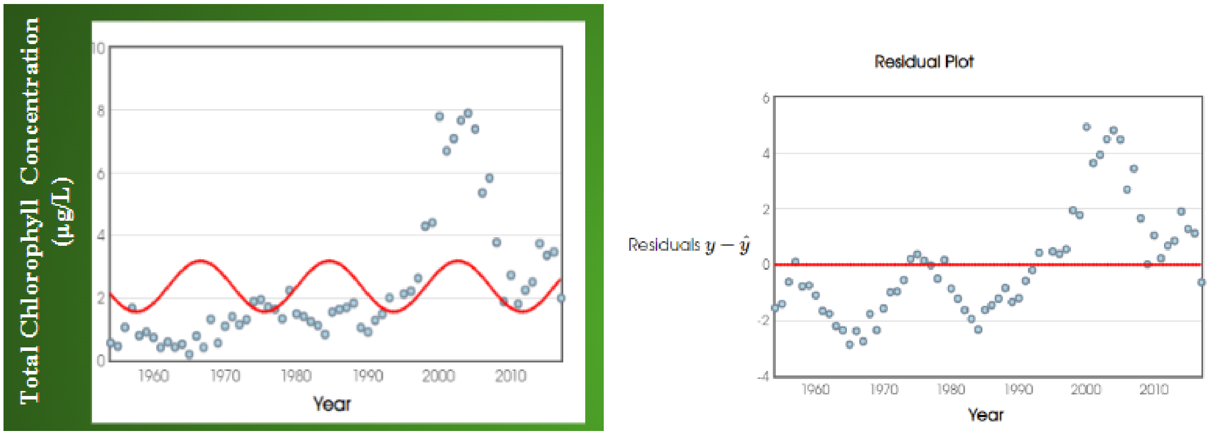


Regression Equation: $\mathrm{TEMPERATURE} = 1.3548 \cdot \sin(0.2327\mathrm{YEAR} + 1.7855) + 9.5894$

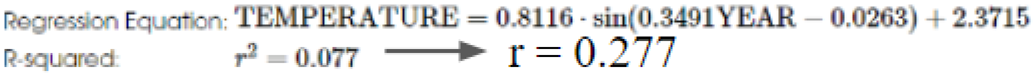
R-squared: $r^2 = 0.3868 \longrightarrow \mathrm{r} = 0.6219$

Figure 2

Time Series of Average Global Oceanic Chlorophyll a Concentrations (µg/L) with Residual Plot



Note. Preliminary sinusoid for average total oceanic chlorophyll with residual plot.

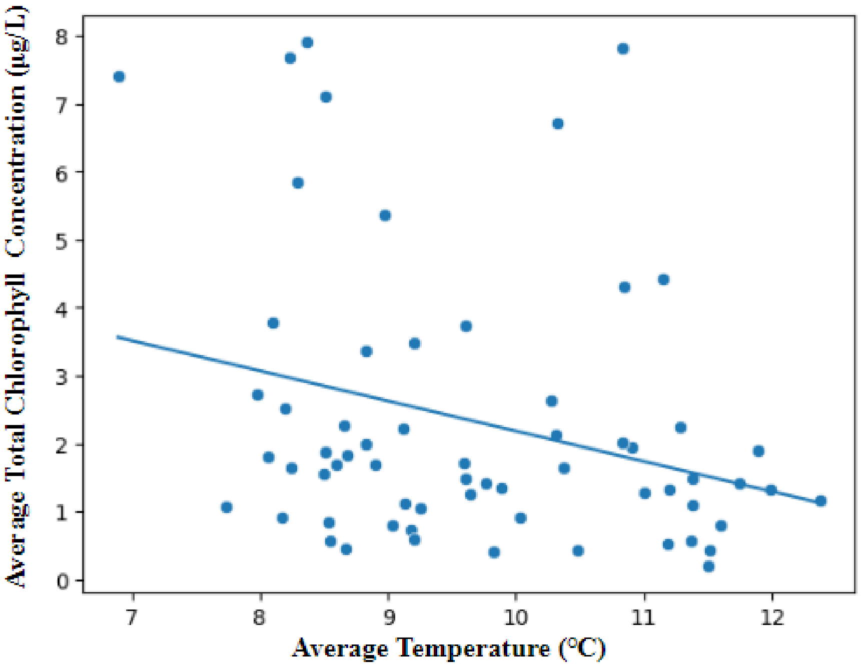


Regression Equation: $\mathrm{TEMPERATURE} = 0.8116 \cdot \sin(0.3491\mathrm{YEAR} - 0.0263) + 2.3715$

R-squared: $r^2 = 0.077 \longrightarrow \mathrm{r} = 0.277$

Figure 5

Temporal Linear Regression of Average Global Chlorophyll Concentrations (µg/L) Given Average Global Temperature (℃)



Note. Preliminary linear regression between average oceanic temperature and average total oceanic chlorophyll. $y = -0.4438x + 6.6180$; $R^2 = 0.08$ ($r = 0.2828$) $p = 0.0235$* ($\alpha < 0.05$)

Calculations and Data Analysis:
Average Global Oceanic Temperature Time Series:
-Midline: 9.5894℃ | Amplitude: 1.3548℃ → Temperature Fluctuation Between 8.2346℃ and 10.9442℃
-High Outliers from 1916-1918 (small sample sizes)
-Slight Downward Concavity in Residuals Plot

Average Global Total Oceanic Chlorophyll Time Series:
-Midline: 2.3715 µg/L | Amplitude: 0.8116 µg/L → Total chlorophyll Fluctuation Between 1.5599 µg/L and 3.1831 µg/L
-Sudden Increase in Concentrations During Early 2000s
-Geospatial attributes as a possible confounding factor
-Multiple Oscillations of Varying Amplitude in Residuals Plot

Linear Regression Between Variables:
-The high chl a values from the 2000s contribute to the linear model shown
-These outliers occur for non-extreme temperature, only lowering R2, but not p
-This suggests that the model, barring extreme scenarios, can provide predictions for chl a levels to a reasonably accurate degree
-Negative correlation between temperature and chlorophyll concentrations
-Predicted decrease of ~0.4438 µg/L in chl a for every increase in 1℃ temperature
-Consequently, more broadly, this relationship suggests a decline in phytoplankton primary production and biomass as global sea temperatures rise due to global warming.

From these three regression models, it is clear that there is limited model fitness, although important ramifications have been provided nonetheless.

Concluding Remarks:

This preliminary trial of the final model apparatus has proven to be viable in producing somewhat accurate results, though accuracy could be improved upon. More importantly, it is clear that environmental and climatic ramifications can be derived. Therefore, the next steps are to iterate this process upon cleaner data such that model fitness is increased. Some of the insights on temperature may be suitable for future research. Given that the OSD is at a highly global scale, the applicability of the results, when derived from this methodology, will surely make a project that produces sound and significant results.

2324_STEMProject_Exp2_Data_Sharma_v1
2324_STEMProject_Exp2_Code_Sharma_v1

Experiment 4:
Development of Sinusoids for the Time Series
25 January 2024
Introduction:

In order to establish overall model validity as well as gauge forecasting abilities, sinusoidal curves of every parameter were created. Having time series models to predict parametric values would be highly useful for policymakers and scientists alike, concerned with looking at local aquatic ecosystems. Here, sinusoidal regression for this end is applied at a global scale as a test case.

Methods/Materials:

Environmental features tend to be periodic in nature. The sine and cosine functions provide an effective way to model cyclical trends. Therefore, this specific type of a regression model was chosen, with R^2 and r measuring model strength and accuracy. This allowed for the evaluation of the validity of the overall computational system as well as projection abilities. Equation 1 represents the template function used for all time series models:

$$f(\kappa) = Asin((2\pi\gamma)\beta + \varepsilon) + \phi \tag{1}$$

Where $f(\kappa)$ is the function for the total chlorophyll concentration κ, A is the amplitude, 2πγ represents the length of the period (using radians, γ alone being in degrees), β is the given environmental factor (the next section enumerates the variable designation of each parameter), ε is the phase shift, and ϕ is the offset.

Additionally, using the offset as a midline for the sinusoid and the amplitude as a sort of ruler, an interval of all measurements projected by the sinusoid of each parameter was developed. Equation 3 represents the basic construction of the described sinusoidal interval:

$$\phi \pm A \tag{2}$$

Observations and Experimental Data:

Table 1

Properties of Sinusoidal Time Series Models of Environmental Parameters

Parameter	Variable Designation	R^2 Value (r)	Offset (ϕ)	Amplitude (A)	Parametric Range Projected by Sinusoid (ϕ ± A)
Salinity (ppt)	s	0.847 (0.920)	30.112	3.724	(26.388, 33.836)
pH	φ	0.825 (0.908)	8.001	0.135	(7.866, 8.136)
Total Chlorophyll (μg/L)	κ	0.648 (0.805)	289.542	289.457	(0.085, 578.999)
Dissolved Oxygen (μmol/kg)	d	0.500 (0.707)	215.281	-12.543*	(202.738, 227.824)
Nitrate (μmol/kg)	η	0.381 (0.617)	-3650.602	3665.327	(-7315.929, 14.725)
Phosphate	q	0.336 (0.580)	1.191	0.118	(1.073, 1.309)

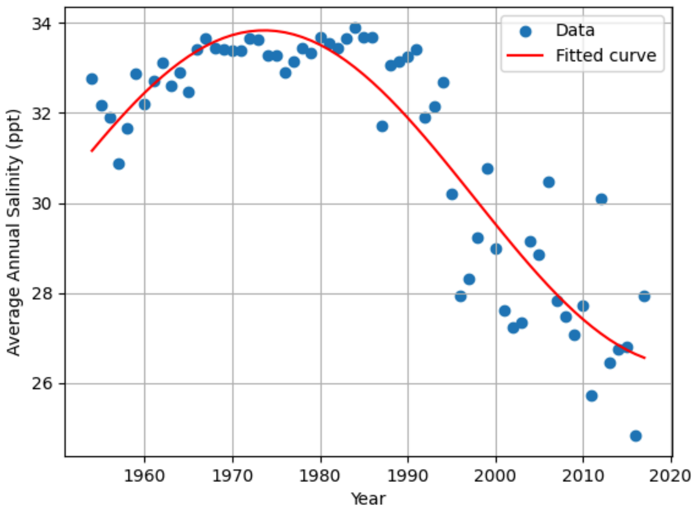
(µmol/kg)					
Temperature (℃)	t	0.327 (0.572)	9.640	1.045	(8.595, 10.685)
Silicate (µmol/kg)	h	0.270 (0.520)	33.095	6.120	(26.975, 39.215)
Alkalinity (milli-equivalent/liter CaCO3)	c	0.239 (0.489)	2.268	0.098	(2.170, 2.366)
Pressure (decibars)	ρ	0.077 (0.277)	489.463	109.099	(271.265, 598.562)

Note. *Although the amplitude for the sinusoid projecting dissolved oxygen is negative, this does not interfere with interval construction, as that is a matter of both adding and subtracting the magnitude of the amplitude from the offset, meaning the net output is the same. In addition to the ones listed above, variable "y" represents the year. The denoted variables for each parameter are to be used in all future logbook entries.

R^2 values of these regression models measure the proportion of the variation observed in the model predictions that are a result of the actual oceanographic data collected. This means a higher R^2 value indicates greater strength in model forecasting capabilities. In conjunction with this table, the sinusoidal regression for salinity, the parameter with the highest R^2 value, as well as pressure, the parameter with the lowest R^2 value, are provided. By representing the most and least fit models, the sinusoidal regression models for salinity and water pressure act as landmarks for the upper and lower bounds in forecasting capabilities observed among the time series models produced. Time series models of other parameters can be found in the file pasted below.

Figure 1

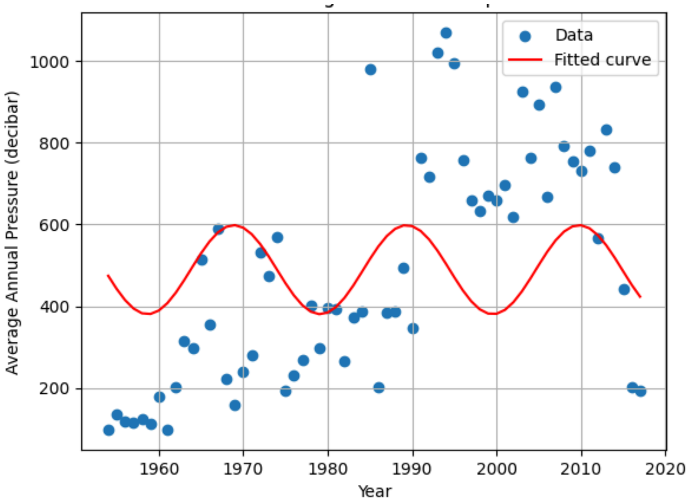
Time Series of Average Global Oceanic Salinity (ppt) from 1954-2017



Note. The time series for salinity (s) for year y is described by the sinusoidal regression function of f(s) = 3.724*sin(6.220y + 80.552) + 30.112; R^2 = 0.847. Blue points represent individual average salinity levels, while the red line depicts the sinusoid.

Figure 2

Time Series of Average Global Oceanic Pressure from 1954-2017 (Decibars)

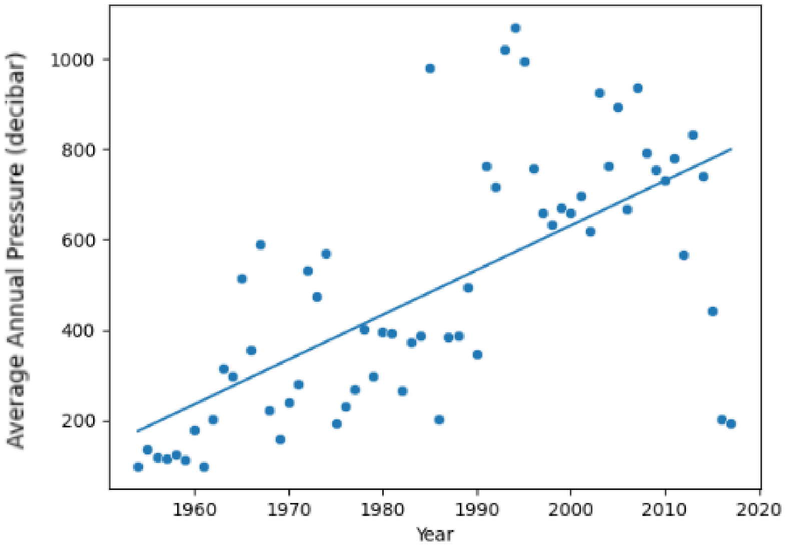


Note. The time series for pressure (ρ) for year y is described by the sinusoidal regression function of f(ρ) = 109.099*sin(6.951y + 43.661) + 489.463; $R^2 = 0.077$

Due to the extremely low R^2 value observed for water pressure, to increase forecasting capabilities, an alternative time series was constructed using linear regression. This model had a higher R^2 value and demonstrated a significant increase in water pressure from 1954 to 2017 (Figure 7, $\alpha = 0.05$, $p < 0.0001$***).

Figure 3

Alternative Time Series of Average Global Oceanic Pressure from 1954-2017 (Decibars)



Note. The alternative time series for pressure (ρ) for year y uses linear regression is described by the function of f(ρ) = 9.904*y - 19176.54; Since R^2 = 0.44, the forecasting fitness for water pressure has successfully been increased.

Analysis and Conclusions:

Given the wide range of R^2 values, in conjunction with the varying practicality of the quantities projected by the sinusoidal intervals observed among all environmental variables, overall model fitness appears to be moderate, with extreme variability given the parameter of interest.

Salinity has the highest R^2 value, standing at 0.847 (Table 1). In context, this suggests that about 84.7% of the variability seen in model predictions of salinity is as a result of the relationship between predictions and the temporal progression of oceanic salinity. This strong connection is visually complemented by the relative proximity of data points to the sine wave (Figure 1). This provides evidence that some parameters can in fact be reliably forecasted using a computational time series. On the other hand, water pressure has the lowest R^2 value, sitting at 0.077 (Table 1). This means that only approximately 7.7% of model predictions are a result of the relationship held with the temporal distribution of water pressure. The weak connection between the variables is made apparent by the divergence of the majority of data from the projected sinusoid (Figure 2). However, it is clear that this lack of fitness can be rectified by using an alternative function. In the case of pressure, using linear regression as opposed to sinusoidal regression increases model fitness, with R^2 rising to 0.44 (Figure 3). Pressure acts as a parameter that provides conclusions contradictory to those of salinity. Whereas salinity's results indicate the possibility of a time series to accurately model parametric values, pressure's results provide evidence of the existence of cases where the exact opposite is true. Moreover, given that an improvement of model fit was attained via using an alternative form of regression, the necessity of employing multiple types of functions to maximize model fitness, as opposed to homogeneously using one function as done in this paper, is made clear. Salinity and pressure merely represent the upper and lower bounds of forecasting capabilities. Between R^2 = 0.847 and R^2 = 0.077 lie a range of R^2 values that represent varying abilities to provide accurate forecasting as well as varying degrees of divergence of data from the centralized sinusoid.

Another indication of model fitness is the sinusoidal intervals calculated by using the offset and amplitude values of the sinusoids. In other words, the values of the peaks and troughs of the sine wave were noted. In some instances, these values aligned very closely with the data. For example, salinity has a peak-trough interval ranging from 26.388 ppt to 33.836 ppt, which encompasses the range of the majority of data (Table 1; Figure 1). By contrast, other peak-trough intervals are impractical, including negative values in contexts that do not make sense, as well as going far beyond the range of the values of the empirics being attained and modeled. For example, the troughs of nitrate concentrations reach far into negative values, trough reaching -7315.929 µmol/kg (Table 1). Concentration levels of a substance cannot be expressed as negative values, making these predictions impractical. This limits the temporal applicability of the sinusoidal model, as certain years, when plugged into the model, would result in these quantities. Once more, a gradient among model attributes can be observed, in this case of the practicality of each parameter's sinusoidal intervals. Similar model liabilities appear to be present with total chlorophyll concentrations, whereas other parameters, such as phosphate and temperature, have sine waves whose peaks and troughs properly encompass experimental values.

2324_STEMProject_FinalCumulativeExp_Data_Sharma_v1
2324_STEMProject_SineReg_Code_Sharma_v1
Supplementary File (includes sinusoids of all parameters, namely the ones not shown above)

Experiment 5:
Linear Regression Models of Parametric Relationships with Total Oceanic Chlorophyll
27 January 2024
Introduction:

In order to assess direct prediction capabilities of individual parameters for total oceanic chlorophyll, as well as the directionality of relationships, linear regression models of average total oceanic chlorophyll given each parameter were created. The directionality of relationships could have significant ecological and climatic ramifications. For example, if temperature is found to have a positive correlation with e.g. biomass, as an aquatic system warms, it would be predicted that

Methods/Materials:

For the relationship of every parameter with the indicator, total chlorophyll concentration, a Linear Regression model was developed. On a functional level, assessing each individual parameter's relationship with total chlorophyll acted as a precursor to identifying which of them had a significant influence on chlorophyll when all parameters were considered in tandem. In essence, performing linear regression acted as a prerequisite for then performing PCA. The significance of relationships were determined using a Student's t-test for Linear Regression at $\alpha = 0.05$. Both double- and single- tailed p-values were attained. This was done in conjunction with the use of R^2 and r to measure model accuracy and strength. Similar to the previous set of models, Equation 1 provides a template wherein each parameter's regression model was represented:

$$f(\kappa) = m\beta + \beta_0 \tag{1}$$

Where $f(\kappa)$ is the function for the total chlorophyll concentration κ, m is the predicted slope of the line, β is the given parameter, and β_0 is the y-intercept of the model.

Observations and Experimental Data:

Table 1

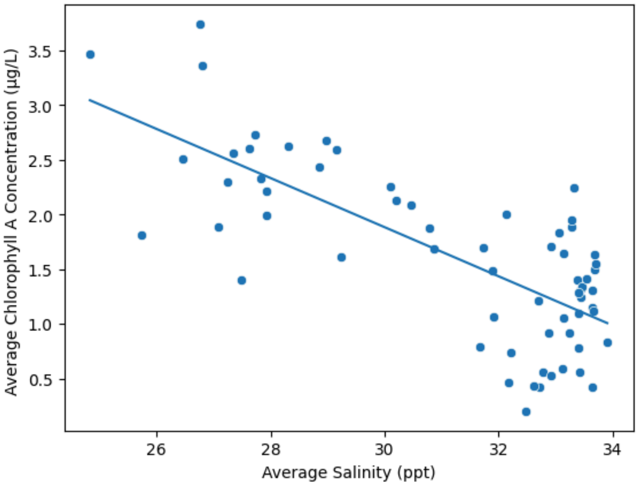
Linear Regression Information of Parametric Factors Related to Total Oceanic Chlorophyll

Variable	R^2 Value (r)	p-value (single-tailed p-value) $\alpha = 0.05$	Equation
s	0.54 (0.735)	0.0000 (0.0000)***	$f(\kappa) = -0.2248s + 8.6275$
φ	0.27 (0.520)	0.0000 (0.0000)***	$f(\kappa) = -3.9346\varphi + 33.0813$
d	0.25 (0.500)	0.0000 (0.0000)***	$f(\kappa) = 0.0333d - 5.6467$
ρ	0.22 (0.469)	0.0001 (0.00005)***	$f(\kappa) = 0.0014\rho + 0.9536$
q	0.10 (0.316)	0.0130 (0.0065)*	$f(\kappa) = 1.8526q - 0.5389$
η	0.09 (0.300)	0.0171 (0.0086)*	$f(\kappa) = 0.0634\eta + 0.8126$
h	0.08 (0.283)	0.0244 (0.0122)*	$f(\kappa) = 0.0324h + 0.6060$
t	0.07 (0.265)	0.0318 (0.0159)*	$f(\kappa) = -0.1641t + 3.2049$
c	0.07 (0.265)	0.0368 (0.0184)*	$f(\kappa) = 1.4551c - 1.6810$

Note. From highest to lowest R^2 value, the parameters are: salinity, pH, dissolved oxygen, pressure, phosphate, nitrate, silicate, temperature, and alkalinity. Significance levels of $p \leq 0.001$ are denoted with three astrices. When $p \leq 0.05$ is true, only one asterisk is used. No significance values between 0.01 and 0.001 were attained, so no significance values were denoted with two asterisk. Alkalinity is placed below temperature due to having a lower p-value. This addresses their equivalent R^2 values.

Figure 2

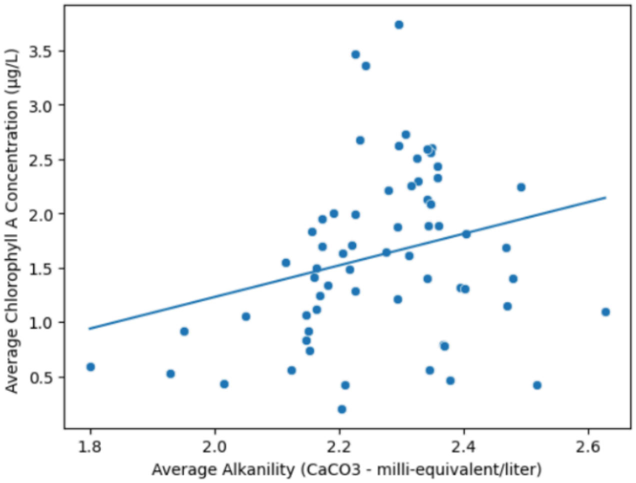
Linear Regression Model of Total Oceanic Chlorophyll (µg/L) Given Salinity (ppt)



Note. This regression model, with the highest R^2 value, possesses the strongest predictive capability between total chlorophyll and an environmental measure. Blue points represent corresponding salinity and chlorophyll measurements for each year between 1954 and 2017, while the equation, f(κ) = -0.2248s + 8.6275, is represented by the blue line.

Figure 3

Linear Regression Model of Total Oceanic Chlorophyll (μg/L) Given Alkalinity (meq/L of CaCO$_3$)



Note. This regression model, with the lowest R^2 value, possesses the weakest predictive capability between total chlorophyll and an environmental measure. Blue points represent corresponding alkalinity and chlorophyll measurements for each year between 1954 and 2017, while the equation, f(κ) = -0.2248s + 8.6275, is represented by the blue line.

Analysis and Conclusions:

An analysis of R^2 and p values among the series of linear regression models depicting the relationship of different parametric variables with chlorophyll concentrations reveals that while relationships are weak, the directional value of each relationship is statistically significant.

The highest linear regression model had an R^2 value of about 0.54, that being for the relationship of chlorophyll levels given salinity, whereas all other models have ones below 0.3 (Table 1). This means that for most of the time, less than 30% of the observed variation in chlorophyll concentrations is due to its relationship with the given environmental variable. Looking at scatterplots for both the strongest and weakest relationships, most data diverge significantly from the trendline (Figure 2; Figure 3). Nonetheless, the directionality of these relationships are still statistically significant, with the p-values for all double-tailed t-tests for linear regression being less than $\alpha = 0.05$. The single-tailed p-values, being half the amount and focused specifically on relationship directionality, provide even stronger evidence for observed positive and negative relationships between chlorophyll and parameters (Table 2, $p < 0.05$*). By seeing how each individual parameter impacts chlorophyll concentrations, crucial ecological and climatic insights can be drawn, given the pigment's role in facilitating photosynthesis, which in turn influences the transfer of trophic energy, sequestration of carbon, cycling of biogeochemical nutrients, and other important functions for the global climate and environmental systems.

Chlorophyll concentrations hold a negative relationship with pH, temperature, and salinity (Table 1, $p < 0.05$*). Indeed, past literature has noted the overall increase in global oceanic temperatures and acidity (Berwyn, 2018). Moreover, salinity is known to inhibit chloroplast activity (Hnilickova et al., 2021). The implications of rising temperature, acidity, and salinity are not simple directional impacts on phytoplankton norms. Primary production capabilities (and more broadly, other traits), increase alongside temperature, pH, as well as any other parameter, until the optimum level is reached, after which there is a decline (Dedman et al., 2023). Using chlorophyll concentrations as an intermediate indicator, this may imply that many phytoplankton species are under conditions suboptimal for optimally performing primary production. This may be observable in the form of slower metabolic rates and other biological indicators. However, lower chlorophyll concentrations would indicate lower metabolic capabilities for phytoplankton, due to the implied dearth of resources to photosynthesize. These data provide evidence that as ocean temperatures warm, primary production in phytoplankton shall decline. This means lower levels of energy being sent up the trophic pyramid, lower rates of nutrient cycling, and the inhibition of carbon sequestration and other climate regulation processes. In addition to this overall decline, primary production levels can be expected to become increasingly heterogeneous along the spatial gradient. Salinity and nutrient concentrations are projected to become less uniformly concentrated (Berwyn, 2018). With the former parameter holding a negative relationship with total chlorophyll (Table 1), this implies that the decrease in chlorophyll will be dissimilar among locations, provided different salinity levels. A similar logic may be applied to the latter variables, which have positive relationships with total chlorophyll (Table 1). This lack of spatial homogeneity in primary production further limits ecological stability. Therefore, as oceanic parameters continue to evolve, it appears that the stability and health of climatic and ecological systems shall continue to decline.

Meanwhile, all other parameters, mainly including oxygen and various micronutrients, hold a significant positive relationship with chlorophyll concentrations (Table 1, $p < 0.05$*). Indeed, when there is a greater presence of chlorophyll, that indicates that a greater amount of photosynthesis can occur, stimulating subsequent metabolic pathways that facilitate nutrient cycling, allowing for micronutrient concentrations to grow. This tie of chlorophyll to the stimulation of micronutrients could have led to the positive relationships observed. In essence, the data may serve as support for principles in biogeochemical cycling as well as similar areas.

These results indicate the need to address ways in which rises in ocean temperature can be perturbed, as well as the need to regulate the concentrations of nutrients and other variables. In conjunction with the time series models provided above, this provides a potent source for prediction and decision-making. Knowing the impact a given level of a parameter may have on an aquatic ecosystem can be crucial for policymakers and scientists. If it is known, from a strong time series model, that a certain level of, for example, phosphorus, shall lead to a harmful amount of eutrophication (or other phytoplankton trait), a conclusion reached from observing a model from the current set being presented, then decisive policy action may taken, provided these crucial details.

2324_STEMProject_FinalCumulativeExp_Data_Sharma_v1
2324_STEMProject_LinReg_Code_Sharma_v1
Supplementary File (Includes all the linear regression relationships not shown above.)

Experiment 6:
Principal Component Analysis and Covariance Matrices
29 January 2024
Introduction:

In order to determine driving parameters and inter-parameter relationships, PCA was performed, and correlation matrices were developed. It is important to note that driving parameters and inter-parameter relationships vary from ecosystem to ecosystem. The temporal range being observed also plays a role in this. For example, only 1954-2017 data of global ocean conditions were observed. Any different set of spatiotemporal conditions would undoubtedly reveal different results. Nonetheless, observing these characteristics at a global scale provides a good example of how this methodology can be applied. It can also provide information on these specific metrics for the global environment, offering a crucial view of marine ecosystems at the global scale.

Methods/Materials:

In order to identify the driving parameters behind total chlorophyll concentrations, PCA was used. PCA is a dimension reduction technique that compresses multiple independent variables into fewer dimensions so as to summarize overarching data patterns and allow for ease of data visualization. Before performing PCA, the data must be standardized so that scale does not impede the accuracy of results. In this study, before PCA was performed, all data of the indicator (chlorophyll) and every parameter were standardized using minimum-maximum normalization. Within the setting of PCA, the total variance of the data is measured. This variance is captured by a finite set of portions of the data known as principal components. In two-dimensional representations, the first two principal components, that is, the two components that account for the highest amount of variance, denoted PC_1 and PC_2, are placed on the horizontal and vertical axes respectively. Since information from the other principal components (PC_3, PC_4, … PC_n) is omitted, it is important that most variance is captured by PC_1 and PC_2. This is measured by each principal component's eigenvector values, which are derived from various matrix operations performed on the data. Then, to standardize the amount of variation each principal component captures, the eigenvalues are divided by the total variance of the dataset. A scree plot is used to depict the cumulative coverage of variance by all principal components. Along with a PCA plot, a scree plot was used to illustrate these notable properties of the principal components. In two-dimensional PCA, every parameter assessed captures some amount of either principal component, and holds either a positive or negative relationship with the directionality of component variances. This magnitude and directionality is represented by a pair of coordinates that form a vector. The greater the magnitude of variance represented by a parametric vector, the more influential that parameter is relative to the overall data, and in turn, driving the dependent variable. For the study, the magnitude of each parameter was calculated as depicted by Equation 1:

$$M(\beta) = \sqrt{(C_{PC1})^2 \cdot \nu_{PC1} + (C_{PC2})^2 \cdot \nu_{PC2}} \tag{1}$$

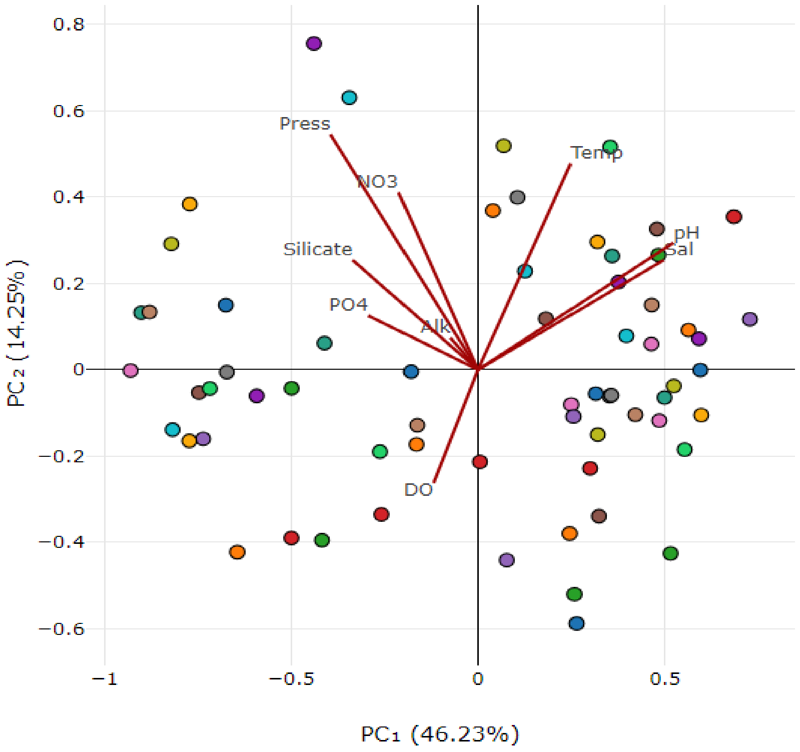
Where $M(\beta)$ is the function of the magnitude of parameter β, C_{PC1} is the contribution of the parameter to the variance of PC_1, while C_{PC2} is the contribution of the parameter to the variance of PC_2, and ν_{PC1} is the proportion of the total variance represented by PC_1, and ν_{PC2} is the proportion of the total variance represented by PC_2.

The magnitudes of each parameter were calculated using the above equation, and then subsequently ranked by descending magnitude values. The parameters with the highest calculated magnitude were identified as driving parameters of chlorophyll concentrations. Additionally, to assess the presence of inter-parameter relationships, a covariance matrix was used. A covariance matrix is an intermediate operation performed in the complex matrix calculations involved with PCA wherein all independent variables are arranged in a square array. The cells of this matrix contain the covariance between the row and column parameters. Values vary between 0 and 1, and can be either positive or negative based on the directionality of the relationship. A greater magnitude indicates a stronger relationship between the two parameters. The diagonal cells represent the variance of that individual parameter following dimension reduction procedures.

Observations and Experimental Data:

Figure 1

PCA Plot of Driving Parameters Behind Total Oceanic Chlorophyll Chlorophyll Concentrations

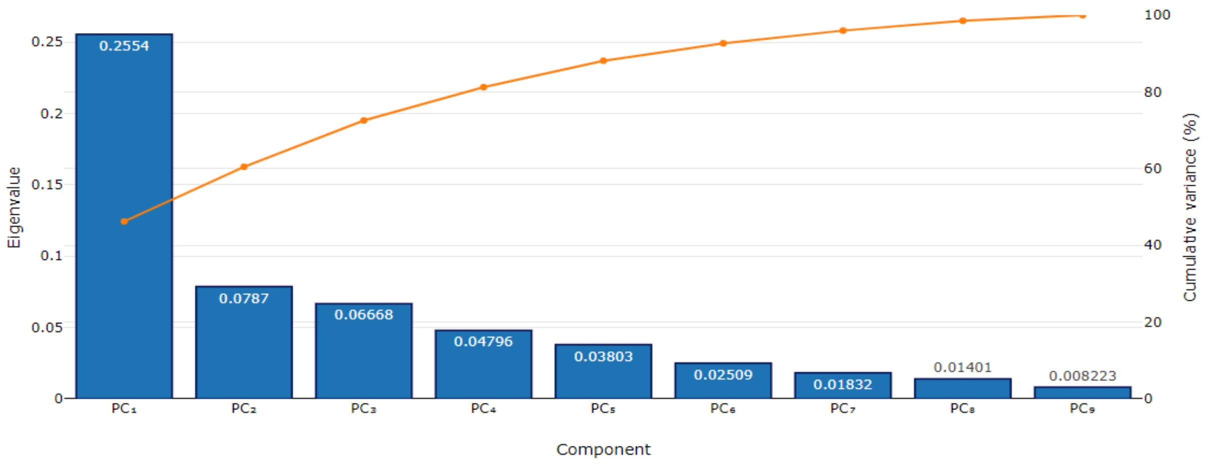


Note. The horizontal and vertical axes are represented by PC_1 and PC_2 respectively. The individual dots of varying color represent various instances of location of chlorophyll measurements relative to the first two principal components following dimension reduction. The red lines sprouting from the origin represent the vectors of each parameter's contribution to the variance of the first two principal components. Parametric abbreviations are designated as follows: pH (pH), salinity (sal), temperature (temp), nitrate (NO_3), pressure (Press), silicate (Silicate), phosphate (PO_4), alkalinity (alk), and dissolved oxygen (DO).

PC_1 accounts for 46.23% of overall variance of the dataset, whereas PC_2 accounts for 14.25%. Figure 2 provides a scree plot depicting the cumulative coverage of whole-dataset variance among the total nine principal components.

Figure 2

Scree Plot of Cumulative Dataset Variance Coverage Among Principal Components



Note. The blue bars with the white numbers represent the contribution of each principal component to total variance in the form of an eigenvector value. This is not to be confused with the standardized contribution of each component to total dataset variance, which is attained only after dividing these eigenvector values by the total dataset variance. The orange line represents the cumulative progression of coverage of this variance.

As with the previous two stages of computational models, a table utilizing a color gradient to represent the progression of statistical measurements is provided. Table 1 orders parameters from the highest to lowest magnitude, the measurements of which were based off of the operations enumerated in Equation 1.

Table 1

Magnitude of Parametric Contributions to the Primacy Principal Components

Variable	Contribution to PC_1	Contribution to PC_2	PC_1/PC_2 Magnitude*
φ	0.5229155	0.2942522	0.3724914
s	0.4967389	0.2542255	0.3511156
ρ	-0.3944697	0.5453883	0.3381172
h	-0.3350876	0.2544652	0.2472569
t	0.2481767	0.4774918	0.2469081

η	-0.2119327	0.4106519	0.2116481
q	-0.2929772	0.1256602	0.2047729
d	-0.1187105	-0.2619915	0.1276556
c	-0.07405966	0.07430118	0.0576397

Note. *As enumerated in equation for 4, these measurements were attained via the following formula (refer to Equation 4 for variable definitions): $M(\beta) = \sqrt{(C_{PC1})^2 \cdot \nu_{PC1} + (C_{PC2})^2 \cdot \nu_{PC2}}$. Given that all contributions are squared, regardless of the directionality of the contribution of parameters towards the first two principal components, the magnitude is expressed as a positive value.

Finally, the last data analysis technique presented is a covariance matrix depicting the inter-parameter relationships between the environmental factors within the WOD18 dataset that were tested for this study (Table 2).

Table 2

Covariance Among Oceanic Parameters in OSD Dataset of WOD18, 1954-2017

	Temperature (℃)	Salinity (ppt)	Dissolved O_2 (μmol/kg)	Pressure (decibars)	pH	Alkalinity (meq/L $CaCO_3$)	NO_3 (μmol/kg)	PO_4 (μmol/kg)	Silicate (μmol/kg)
Temperature (℃)	0.05679								
Salinity (ppt)	0.03166	0.08317							
Dissolved O_2 (μmol/kg)	-0.01647	-0.01959	0.04698						
Pressure (decibars)	-0.005877	-0.03823	0.01677	0.08013					
pH	0.03188	0.06853	-0.01324	-0.03891	0.09541				
Alkalinity (meq/L $CaCO_3$)	0.006023	-0.01187	0.001403	0.009997	-0.01088	0.02999			
NO_3 (μmol/kg)	-0.001303	-0.01835	0.003703	0.03955	-0.01663	0.0009591	0.03353		
PO_4 (μmol/kg)	-0.0178	-0.02756	-0.003317	0.02564	-0.03468	0.002638	0.017	0.05649	
Silicate (μmol/kg)	-0.02261	-0.03465	-0.007556	0.029	-0.03073	0.004527	0.02327	0.02543	0.06992

Note. Green cells represent the diagonal cells of the covariance matrix. These quantities, rather than inter-parameter covariance, represent the variance observable within the specified parameter. For example, the cell 0.08013 represents the variance seen within water pressure. By contrast, the cell below, -0.03891, represents the covariance between pH and water pressure.

Analysis and Conclusions:

Based off of PCA results, as well as data provided the covariance matrix and scree plot, it is apparent that pH, followed by salinity and pressure, on a global scale, are driving parameters behind chlorophyll concentrations, and in turn, primary production capabilities. Moreover, each parameter is independent of one another. However, these conclusions are limited by the low variance coverages of PC_1 and PC_2. Being a two-dimensional PCA, much variance information was lost by solely focusing on PC_1 and PC_2. Together, these contain an eigenvector value of only about 0.33 (Figure 2). When standardized, this means that only about 60% of variance of the overall dataset is covered (Figure 1). With nearly half of the data information lost from the process, the conclusions that can be drawn have a rather limited scope.

The contribution of each parameter to the variance of both PC_1 and PC_2, providing further evidence of inadequate variance coverage. Being measured on a scale with a magnitude of 1, the highest contribution to variance coverage of PC_1, captured by pH, was only 0.523, while the highest contribution to variance coverage of PC_2, captured by water pressure, was just 0.545 (Table 1). These contributions are at best, moderate in coverage. Seeing as most contributions are lower than this value, it is clear that most parameters do not particularly account for overall data variance. Nonetheless, in calculating the magnitudes (Equation 1) and ordering the results, it was found that pH, followed by salinity and pressure, are driving parameters of total oceanic chlorophyll concentrations, and in turn, primary production and other aspects of ecology (Table 1). While this may indicate a potential need to study the impact of these parameters on phytoplankton primary production and other dynamics, such a decision must be made cautiously. This because of both the lack of variance coverage from which these results are drawn, as well as the fact omitting factors would create a less representative understanding of empirical quantities and trends.

The covariance matrix values suggest data homogeneity within parameters and independence among different factors. Among the diagonal cells of the covariance matrix, the highest observable covariance stands at 0.083, with salinity. That means, among the values serving as measures of variance, that is, spread, the highest among these values was only 0.083, on a scale of 1 (Table 2). With all intra-parameter variance values being of this small of a magnitude, it appears that data values for parameters are homogenous. This may provide evidence into the consistency of the data. While this analysis was performed on a global scale using data spanning sixty years, even at scope this broad, some level of parametric homogeneity and consistency of ocean data is implicated. The diagonal cells, consequently, provide indirect evidence for the ocean as a system with properties that have a noticeable level of stability. Moreover, all other cells have even smaller covariance values for the inter-parameter relationships, the vast majority failing to exceed a value of 0.1 (Table 2). With highly weak covariance values, this indicates that oceanic variables may be independent of one another. This is in terms of impacting the levels among one another, rather than with regards to phytoplankton dynamics. Given the focus of this paper and the results compiled among these three sets of computational models, it is clear that parameters exert a complex net impact on phytoplankton dynamics, even if they themselves may not impact their own values.

An additional observation is that salinity, and to an extent, pH and pressure, had some combination of notable R^2, p-values, covariance, and principal component contribution values across all three sets of computational models.

In terms of policy and scientific investigation, a tool such as PCA and covariance matrices would serve as preliminary analysis tools for aquatic ecosystems. This would help establish a framework of general understanding of a given ecosystem. From this vantage point, deeper empirical trends can be performed. Subsequently, this would allow for the development of both scientific and policy-related investigations, allowing for insights to be reached.

2324_STEMProject_FinalCumulativeExp_Data_Sharma_v1

- Min-maxed aggregate.csv contains the standardized data used in forming the PCA plot and covariance matrix.

Supplementary File