headwatersscienceinstitute.org

# HEADWATERS SCIENCE INSTITUTE
## Student Research Journal

Volume 1 • Issue 1 • Fall 2020

Kayla Brand collects water quality data during the summer research program.

# Forward

The articles you see in this publication are the result of months of hard work from students in Headwaters Science Institute's 2020 Summer Research Program. While they are not peer-reviewed in typical journal fashion, they have been reviewed by Headwaters staff and professional scientists. The caliber of these projects speaks loudly to these students' accomplishments. Allow us to also highlight a few of the most special parts of this program here.

Part of what made this program possible was participation by a great array of professional scientists, post-docs, and graduate students from institutions across the country. These scientists not only brought great technical knowledge to the program but also shared with students their own invaluable experience in navigating the winding path of original scientific research. Thank you to all of the scientist mentors who made these student projects possible.

One of the unique outcomes of this course was that all of the articles in this text are the students' own original ideas. Each one of these manuscripts started as part of a list of scientific questions brainstormed by students. From there, students shaped and molded the questions they were interested in based on current scientific knowledge and access to field sites or existing datasets. Students revised their questions, collected preliminary data, and revised again, honing their scientific skills along the way. The diversity of research questions in this publication is not only evidence of the incredible creativity inside all students but also honors the wonderful job the scientist mentors did in nurturing this curiosity.

To the students who authored these papers: congratulations! All of you have worked incredibly hard to complete these manuscripts. More than having your name on this paper, be proud of all that it represents. Each one of you overcame setbacks, roadblocks, time crunches, and uncertainty to get to this point. The skills you practiced along the way will serve you well in whatever you choose to do next.

Spencer Eusden

Program Director - Headwaters Science Institute

**Headwaters Science Institute Journal Staff**

**Editor**
Spencer Eusden
spencer@headwatersscienceinstitute.org

**Graphic Design and Production**
Megan Seifert
meg@headwatersscienceintitute.org

**Assistant Editors**
Le'a Gleason, Daniel Dudek, Jack Holmes

**Mentors**
Jian Zhang, Daniel Dudek, Julia Kelson,

**For any questions please contact**
meg@headwatersscienceinstitute.org
P.O. Box 913, Soda Springs, California 95728
(530) 426-3063

**Headwaters Science Institute Board of Directors**

**President**
Andy Giordano
argiordano@gmail.com

**Secretary**
Michelle Furbershaw
furber@pacbell.net

**Treasurer**
Alec Atkin
aatkin@alum.mit.edu

**Other Board Members**
Jack Holmes, Craig Rowe, Erica Seifert



HEADWATERS SCIENCE INSTITUTE

STUDENT RESEARCH JOURNAL

Volume 1 • Issue 1 • Fall 2020

## TITLE ARTICLE

**4** The impact of timing and magnitude of the El Nino-Southern Oscillation and PDSI levels on the local temperatures in the Bay Area **By Ryan Li**

## ENVIRONMENTAL ANALYSES WITH R

**10** Correlation between shutdown orders and CO levels across the United States **By Arnav Gupta**

**15** Is there a pattern between genetics and future growth/reduction in the ecological niche of a species within the Cypripedium genus? - Insights based on remote genetic data and ecological niche modeling **By Lauren Lee**

**21** Presence of Vegetation in Relation to Slope in Yosemite Valley, California **by Colin Saltzgaber**

**26** Investigating the Relationship between Root Carbon Pools and Soil Organic Carbon Pools **by Sabrina S. Ahmed**

# The impact of timing and magnitude of the El Nino-Southern Oscillation and PDSI levels on the local temperatures in the Bay Area

**Ryan Li[1] and Jian Zhang[2]**

[1] Monta Vista High School, [2] National Oceanic and Atmospheric Administration National Severe Storms Lab

## ABSTRACT

**An analysis of temperature, Multivariate El Niño-Southern Oscillation (ENSO) Index (MEI) and Palmer Drought Severity Index (PDSI) data from the San Francisco Bay Area from 1971 to 2016 was conducted in the current study. $CO_2$ records for Trinidad Head, CA for 2004 to 2016 were also analyzed in the current study along with the annual temperature anomalies for the Bay Area. There was a clear correlation between warm anomalies and high MEI/low PDSI values on an annual basis, however only when both MEI and PDSI were significant. On the three-year moving average MEI had a stronger correlation with annual temperature anomalies than PDSI. Although there wasn't a definitive link between the $CO_2$ of Trinidad, CA and the annual temperature anomalies in the Bay Area, the consistent increase in $CO_2$ concentrations could have had an impact on the increase of annual temperature anomalies from 1971 to 2016, assuming that $CO_2$ concentrations rose at a similar rate from 1971-2003 as well. This study verifies the conclusions that both warm and dry years are happening simultaneously more often as global temperatures increase, which was drawn by previous studies done on the topic of drought and temperature on a regional scale (California), but also looks at another variable, ENSO, and zooms in on a local scale with the study region being the Bay Area.**

## INTRODUCTION

Anthropogenic climate change has over the past 150 years warmed the Earth "at a rate 20-50 times faster" than Earth's fastest natural climate change events, and its effects have been felt around the world (1). This has resulted in not only habitat loss for plants and animals living within ecosystems globally, but also severe infrastructure damage due to more frequent events such as tropical storms, extreme heat events, and disruptions to oceanic circulation which are exacerbated by global warming (2). Yet all of these catastrophic events have occurred when global surface and ocean temperature has only warmed around 0.85°F from 1880 to 2012, while global surface temperature is likely to further rise above 1.5°F relative to 1880 by 2100 with high confidence depicted by models RCP4.5, RCP6.0 and RCP8.5 (3). A major portion of this warming can be directly attributed to human greenhouse gas emissions, specifically $CO_2$, which had an exponential growth as seen in Figure 1.1 (https://www.globalchange.gov/browse/multimedia/global-temperature-and-carbon-dioxide).

The effects of a warming world can also be felt on a regional scale, as is the case with California droughts. Drought is defined as the "temporary reduction in water availability below normal quantities" (4), which can be interpreted as an increase in the dryness of an area. Even though California has always had a long history of drought, Diffenbaugh et al. (2015) stated in their paper that recent years in California had been "very hot and dry at the same time" (5). They go on to say that the probability of this happening in the future would increase, bringing a "double whammy" to the environment with less precipitation and more evaporation/transpiration resulting from warmer temperatures. These conditions can bring about extreme drought events such as the one experienced by California from 2012-2016 (4). Another factor which has a major influence on California's climate is the El Niño Southern Oscillation (ENSO), which is a recurring pattern of oceanic temperatures and atmospheric conditions which have large impacts on rainfall and weather in the United States. This study aims to examine whether the local temperatures of the Bay Area have a relationship to not only $CO_2$ concentrations but also ENSO and California's drought levels over the past 50 years.

## METHODS

The area of study was the San Francisco Bay Area, specifically the three cities of San Francisco, Los Gatos, and Oakland, CA. The time period that was selected spanned from 1971 to 2016, encompassing the period of the 2012-2016 extreme drought in California as well as the forty years preceding the unprecedented warm and dry period. Three weather stations, one from each of these cities, with the most complete temperature records were selected, and monthly average high (Th) and low (Tl) temperatures from each of these three stations were recorded on a spreadsheet. This data was provided by the National Centers for Environmental Information (NCEI) of National Oceanic and Atmospheric Administration (NOAA) through their Climate Data Online tool (https://www.ncdc.noaa.gov/cdo-web/).

The monthly average high and low temperatures reflected the mean fluctuation between the daily high and low temperature in the month for each station. For our study, such small scale fluctuations need to be filtered out to help reveal any long-term trend in the temperature data. Therefore, annual temperature anomalies (Ta) for all three locations were calculated using Equation (1) below:

(1) $$Ta\ (j) = Ty(j)-Tb$$

Here j represents the year and ranges from 1971 to 2016; Ty represents the average annual temperature for each year and Tb represents the 46 year mean temperature. Tb was calculated based on Equation (2) below:

(2) $$Tb = \sum_{j=1971}^{2016} Ty(j)/46$$

And Ty was calculated based on Equation (3):

(3) $$Ty\ (j) = \sum_{i=1}^{12} Tm(i)/12$$

Where i represents the month and ranges from 1 to 12 and Tm represents the average monthly temperature and was defined as follows:

(4) $$Tm\ (i) = [Th(i\ ) + Tl(i)]/2$$

By taking the average of the monthly high and low (Eq.(4)), the diurnal variabilities were removed in the monthly mean temperature (Tm). Then the average of the monthly mean over the year (Eq. (3)) further removed seasonal variabilities in the annual mean temperature (Ty) and subsequently the annual temperature anomaly (Ta). By using the annual temperature anomalies instead of monthly temperature anomalies the potential climate-related trends can be better presented. Thus, the annual temperature anomalies are used for the remainder of this study.

Figure 2.1 shows the annual temperature anomalies from all three stations and similar plots but for each individual station are shown in Fig. 2.2. An example of Los Gatos' annual temperature anomaly bar graph along with its trendline is shown in Fig. 2.2a. The endpoints of each trendline were used to calculate the magnitude of warming for each of the three locations with the equations of the respective trendlines. Then the three warming magnitudes were averaged to represent the average warming in the Bay Area. Afterwards, the annual temperature anomalies for all three stations were averaged to get the Bay Area annual temperature anomaly, which was more representative of the Bay Area's regional climate. In Figure 2.1, 2014 was seen to have a particularly large positive anomaly across all three stations. A paired-t test was run on the average monthly high/low temperatures in all three locations for 2014 against the 1971-2016 average to determine the significance of the temperature increase (Table 2.1). All the p-values indicated an exceptionally warm year in 2014 with high confidence levels.

Next, monthly average $CO_2$ concentration (ppm) from Trinidad Head, CA, which spanned from 2004-2016, was imported onto a spreadsheet. This data was provided by the NOAA Earth Science Research Laboratory (ESRL, https://www.esrl.noaa.gov/gmd/dv/data/index.php?category=Greenhouse%2BGases&parameter_name=Carbon%2BDioxide&search=trinidad%20head&frequency=Monthly%2BAverages) and Trinidad Head was the closest to the Bay Area among all available stations. The annual $CO_2$ anomalies (Ca) were then calculated with a baseline of 2004-2016 using Equation (5):

(5) $$Ca = Cy(j)-Cb$$

Where Cy represents the annual average $CO_2$ concentration while j represents the year that ranges from 2004 to 2016, and Cb represents the 13 year average and was calculated as follows:

(6) $$Cb = \sum_{j=2004}^{2016} Cy(j)/13$$

And Cy was calculated in Equation (7):

(7) $$Cy\ (j) = \sum_{i=1}^{12} Cm(i)/12$$

Where i represents the month and ranges from 1-12 and Cm the monthly average $CO_2$ concentration. This was then placed into a continuous graph with the Bay Area annual temperature anomaly (Fig. 2.3).

Next, bi-monthly data of the Multivariate El Niño/ Southern Oscillation (ENSO) Index (MEI, 6) for 1971-2016, supplied by the NOAA Physical Sciences Laboratory (PSL, https://www.psl.noaa.gov/enso/mei.old/), was placed on another spreadsheet. The MEI is an index that combines several key variables over the tropical Pacific basin (30°S-30°N and 100°E-70°W) that reflect the ENSO phenomenon. The variables include the sea surface pressure, sea surface temperature, zonal and meridional components of the surface winds, and cloudiness of the sky (6). This bi-monthly data was then averaged for each one of these years to obtain the annual mean MEI. Significantly positive (negative) annual mean MEI values 0.5 (-0.5) represent warm El Niño (cold La Niña) events.

Finally, monthly Palmer Drought Severity Index (PDSI, 7) data from 1971-2016 was taken from the Climate.gov's Drought Risk Atlas (https://droughtatlas.unl.edu/Data/Climate.aspx) for each of the three stations, then averaged to obtain the annual mean PDSI for the Bay Area. This ensures that the annual scaling was present throughout all of the datasets. PDSI is an index that measures the magnitude of drought. It has values ranging from -10 to +10 based on the amount of incoming precipitation as well as outgoing evapotranspiration and groundwater recharge (8). Larger positive values indicate a wetter environment while larger negative values indicate the opposite. Then the Bay Area annual temperature anomaly, annual mean MEI, and annual mean PDSI were plotted onto the same continuous graph (Figure 2.4). A 3-year moving average was applied to all three data sets to smooth out some short-term natural variabilities and to help better visualize any long-term trends and relationships in the data.

## RESULTS

The bar graphs for the annual temperature anomalies in Los Gatos, San Francisco, and Oakland each showed a warming trend throughout, with the average warming calculated from the endpoints of their trendlines ranging from 0.8 to 2.6°F (Fig. 2.2). By averaging these values, the average warming for the Bay Area from 1971 to 2016 was found to be 1.4°F, or 0.78°F. Furthermore, the bar graph combining all three locations had a large spike in annual temperature anomaly for all three locations in 2014 ranging from 2.6 to 4.2°F (Fig. 2.1). The paired t-test results (Table 2.1) on the average monthly temperatures in 2014 all had p-values below the 0.05 threshold, rendering the temperatures significant in the historical context.

The continuous graph (Fig. 2.3) plotting both Bay Area annual temperature anomaly and Trinidad Head annual $CO_2$ anomaly showed a steady rise in $CO_2$ levels between 2004 and 2016. There was a significant ($R^2$ > 0.99) trend of $CO_2$ increase at ~1.95ppm/year (Fig. 2.3) even if the data record was relatively short. The Bay Area annual temperature also has a trend of increase at ~0.03 deg/year with a lower $R^2$ = 0.157. For the combined graph (Fig. 2.4) with Bay Area annual temperature anomaly, average annual MEI, and average annual PDSI, the first two variables showed a positive correlation (correlation coefficient = 0.73 between the 3-year moving averages of Ta and MEI), with temperature anomaly trailing MEI until 1997 to 1999, then the MEI began trailing the temperature anomaly after that point. The Bay Area annual temperature anomaly showed somewhat of an inverse correlation with PDSI (correlation coefficient = -0.27 between the 3-year moving averages of Ta and PDSI), becoming the most pronounced in the period after 2000 to 2002. Taking a closer look at the 2013 to 2016 time period, there was a clear spike then gradual decrease of annual temperature anomaly exceeding its fluctuations during the entire

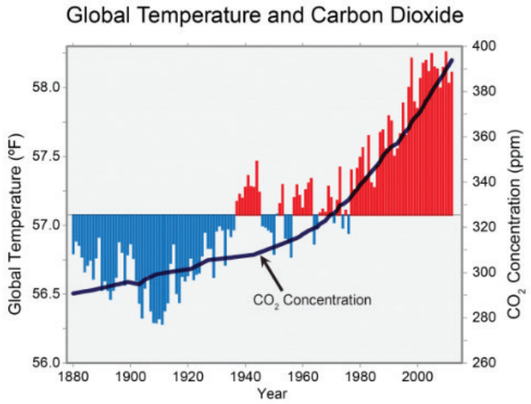



**Figure 1.1** Global temperature and $CO_2$ change over the past 150 years. The bars represent the temperature and black solid line the $CO_2$ concentration.





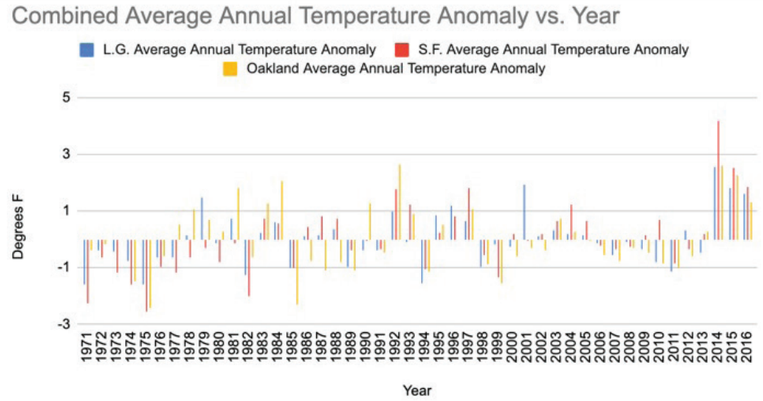



**Figure 2.1** Annual Temperature Anomalies for Los Gatos, San Francisco, and Oakland (1971-2016).

1971 to 2016 period, the PDSI decreased to record lows historically, and the MEI crossed the 0.5 El Niño threshold on a steep ascending trend.

## DISCUSSION

Even though the $CO_2$ concentration in Trinidad Head had a persistent increase between 2004 and 2016 with no particular spike corresponding to the 2014 Bay Area annual temperature anomaly spike, it may still explain the overall increase of the annual average temperature and annual temperature anomaly in the Bay Area during that time period through its role as a greenhouse gas. Assuming that $CO_2$ concentrations also increased at a similar rate from 1971 to 2003 then the overall average increase of 1.4°F in the Bay Area may be attributed to the $CO_2$ induced greenhouse effect. However, it is important to keep in mind that this

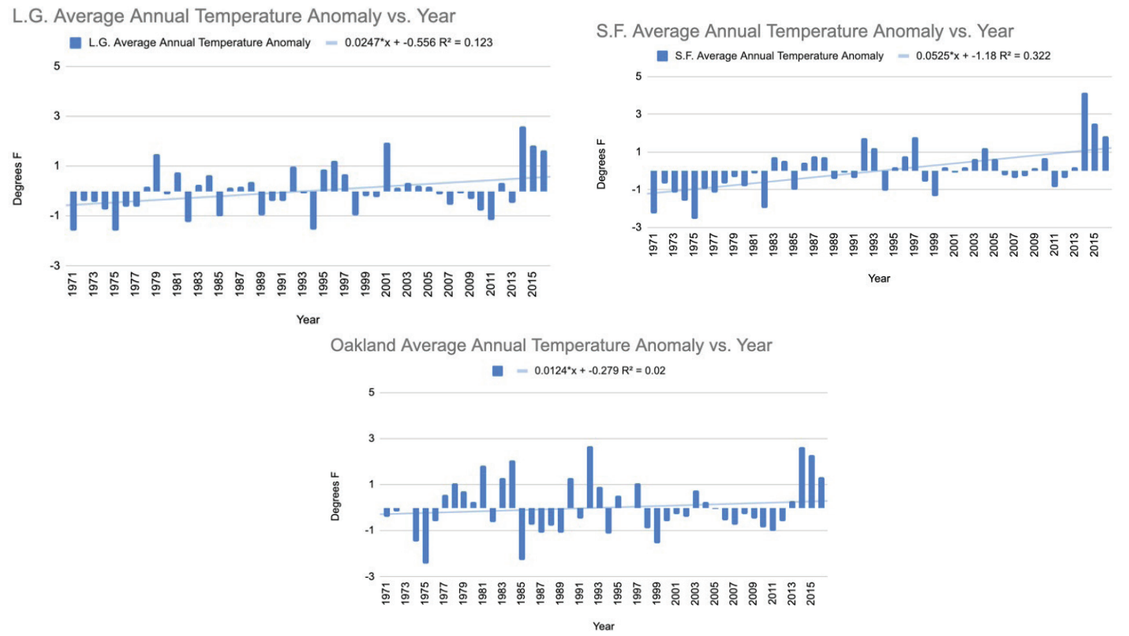



**Figure 2.2** Annual Temperature Anomaly vs. Year For Los Gatos (a), San Francisco (b), and Oakland (c) with Trendlines (1971-2016).

| | Los Gatos (High) | Los Gatos (Low) | SF (High) | SF (Low) | Oakland (High) | Oakland (Low) |
|---|---|---|---|---|---|---|
| Paired t-test results: | 0.0001 | 0.003 | <0.0001 | <0.0001 | 0.0011 | 0.0035 |

**Table 2.1** Paired t-test Results For Average Monthly High/Low vs. 1971-2016 Mean (Los Gatos, San Francisco, and Oakland).

$CO_2$ monitoring station is located around 250 miles from the Bay Area, resulting in the assumption that $CO_2$ trends are similar in different locations. Further study still needs to be conducted on the potential difference in $CO_2$ trends between the Bay Area and Trinidad Head, as well as the spatial variation of $CO_2$ concentrations within California before this assumption can be verified. Furthermore, the exact effect of the $CO_2$ concentration in Trinidad Head on local temperatures in the Bay Area requires further study as well.

During the 2014 temperature spike and the 2013 to 2016 warm period the annual mean MEI crossed over the 0.5 El Niño threshold as the Bay Area annual temperature anomaly jumped above historic levels; however this relationship wasn't significant as the periods from 1981 to 1983 and 1990 to 1993 had annual mean MEI levels at or exceeding 2014 to 2016's levels without a similar spike in annual temperature anomaly. This lack of a large temperature increase was likely due to the fact that there was a record PDSI increase from -1.2 to 4.4 from 1981 to 1983, possibly counteracting the effects of the warm El Niño. However, from 1990 to 1993 the PDSI was always between -1 and 0, indicating that there wasn't much of a wet period to potentially counteract the increasing MEI. Therefore,

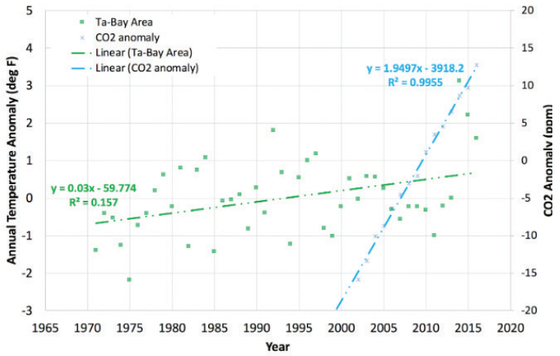



**Figure 2.3** Bay Area Annual Temperature Anomaly (1971-2016, green squares and dot-dashed line) and Trinidad Head, CA Annual CO2 Anomaly (2004-2016, blue stars and dot-dashed line).

a high MEI alone does not necessarily result in large warm temperature anomalies.

The Bay Area annual mean PDSI values from 2014-2016 provided another piece of the puzzle for the temperature spike. During this time the PDSI dropped near or below -2 for four consecutive years starting from 2013 (Fig.2.4), rendering it historically significant. This also shows that it was the driest period over the 46 years studied. This historic drought combined with the following warming El Niño trend may have been a main contributor to the historic temperature anomaly. Other extreme dry periods were seen in 1985, 1988 and 1989 (PDSI = -2.19, -2.03 and -2.09, respectively). However, all three years were in mild to cold La Niña episodes (MEI = -0.39, -0.47, and -0.52, respectively), which might have prevented rapid temperature increases in those years, or just that the MEI values were insignificant and that a low PDSI alone may not cause a significant warm annual temperature anomaly (Fig. 2.4). On the other hand, a cold temperature anomaly in 1975 appeared to be associated with a slightly wet and significantly cold La Niña time period in 1974 (PDSI = 0.47 and MEI = -1.17) and 1975 (PDSI = 0.72 and MEI = -1.28). In general, there seems to be a pattern that severe droughts (wet conditions) concurrent with warm El Niño (cold La Niña) would result in significant warm (cold) anomalies. One exception was around 1996 where a warm annual temperature anomaly (1°F) was associated with a minor wet period (PDSI ~ 1.1) and a weak La Niña (MEI ~ -0.30). This indicates that additional factors may have influenced the temperature record for this year.

Overall, the 2014-2016 historically high Bay Area annual temperature anomaly in conjunction with the historically low Bay Area annual mean PDSI, both of which had simultaneous magnitudes which were unseen of during the 46 year time period, verified Diffenbaugh et al. (2015)'s claim that rising temperatures increase the likelihood of extreme dry conditions and warm temperatures coinciding with each other (5). For the 46 year period analyzed, both the annual mean MEI and PDSI were observed to have a large impact on Bay Area annual temperature anomaly, especially when a high (low) MEI and low (high) PDSI occurred simultaneously. On the other hand, when they were out of sync (i.e. high MEI with high PDSI, low MEI with low PDSI) or only one of them was pronounced, the impact on the annual temperature anomaly can be limited; thus it is shown that both factors need to be significant but also in sync with each other to have the largest influence





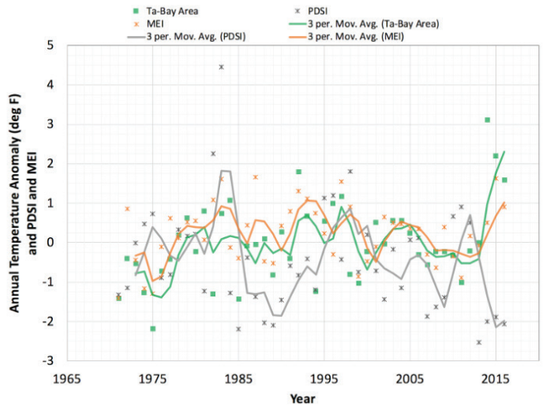



**Figure 2.4** Bay Area Annual Temperature Anomaly (1971-2016, green squares), Annual Mean PDSI (1971-2016, grey stars), and Annual Mean MEI (1971-2016, orange stars) and Their Corresponding 3-year Moving Averages (green, grey, and orange solid lines, respectively).

on the Bay Area annual temperature anomalies. It is noted that while the year to year MEI fluctuations do not necessarily correspond to the year to year annual temperature anomaly fluctuations, with three year moving averages applied, the two had a large correlation (0.73) in general. On the other hand, the PDSI three year moving average had a weak inverse correlation with annual temperature anomaly (-0.27). These results indicate that the Bay Area temperature is significantly influenced by the oceanic system and such influence could be on a multi-year scale, while the relationship between temperature anomaly and PDSI is not as coherent (Fig. 2.4).

Next steps for this research would include incorporating more data series for both temperature and PDSI for additional cities in the Bay Area to ensure that the results are truly representative of the Bay Area, as well as extending the time period examined to further reduce potential bias which may be present in the current study period. The examination of the impacts droughts may impose on Northern California ecosystems would prove to be informative to both the urban ecosystems of the Bay Area as well as the coniferous forests which dominate the Northern Californian landscape.

## ACKNOWLEDGMENTS

I would like to especially thank my mentor Jian Zhang (National Severe Storms Laboratory, NOAA) for providing very insightful advice and edits on my manuscript, as well as guiding me through the research and data analysis process. I would like to thank Spencer Eudsen and Daniel Dudek for creating and running the Monday and Tuesday lectures for the Headwaters Science Institute, as well as being available for advice on designing our projects and writing our research papers during the office hour sessions. I would also like to acknowledge Headwaters Science Institute for providing me with this opportunity to not only meet my knowledgeable mentor but also for me to experience the scientific research process and write my first research paper.

## REFERENCES

1. Nuccitelli, Dana. "New report finds costs of climate change impacts often underestimated." Yale Climate Connections, 18 Nov. 2019, https://www.yaleclimate connections.org/2019/11/new-report-finds-costs-of-climate-change-impacts- often-underestimated/.

2. "The missing economic risks in assessments of climate change impacts." The London School of Economics and Political Science, 20 Sept. 2019, http://www.lse.ac.uk/GranthamInstitute /publication/the-missing-economic-risks-in-assessments-of-climate-change-impacts/.

3. Core Writing Team, Rajendra K. Pachauri and Leo Meyer (eds.). "Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change." IPCC, Geneva, Switzerland, 2014, 151 pp. https://www.ipcc.ch/report/ar5/syr/.

4. Lund, Jay, et al. "Lessons from California's 2012–2016 Drought." J. of Water Resources Planning and Management, Vol. 144, Select0, 2018. https://ascelibrary.org/doi/full/10.1061/%28 ASCE%29WR.1943-5452.0000984.

5. Diffenbaugh, Noah S., et al. "Anthropogenic warming has increased drought risk in California." PNAS, 112 (13): 3931–3936, 2015. https://doi.org/10.1073/pnas.1422385112.

6. Wolter, Klaus, and Michael S. Timlin. "Monitoring ENSO in COADS with a seasonally adjustedprincipal component index." Proc. of the 17th Climate Diagnostics Workshop, Norman, OK, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, 1993, pp. 52-57. Available at: https://www.psl.noaa.gov/enso/mei.old/WT1.pdf.

7. Dai, Aiguo, and National Center for Atmospheric Research Staff (Eds)."The Climate Data Guide: Palmer Drought Severity Index (PDSI)." Climate Data Guide, Last modified 12 Dec 2019. https://climatedataguide.ucar.edu/climate-data/palmer- drought-severity-index-pdsi.

8. Mann, M.E., and Peter H. Gleick. "Climate change and California drought in the 21st century." PNAS, 112 (13): 3858-3859, 2015. https://www.pnas.org/content/112/13/3858.full.

# Correlation Between Shutdown Orders and CO Levels Across the United States

**Arnav Gupta[1] and Daniel Dudek[2]**
[1]American High, [2]University of Texas Arlington

## AABSTRACT

**Research was conducted in summer 2020 to analyze the effect of the shutdown orders due to COVID-19 pandemic on the CO levels across the United States. Since the shutdown orders during the pandemic prevented any public places from operating or conducting any kind of business, the hypothesis was that the CO levels would go down since there was no commute happening or buildings being used. Data was collected from the EPA's Air Quality public data and analyzed in R: a powerful program used to analyze large datasets. The 15 states analyzed were based on the greatest number of cases on August 20th 2020. Each state had its data, with multiple counties and with multiple sites in each county. In almost every single state, the CO levels went down starting from February, with the lowest CO levels during the shutdown, indicating that the shutdown likely could have led to a decrease in CO levels. As some states opened, the CO level started rising then fluctuating.**

## INTRODUCTION

It is no surprise that humans have contributed significantly to climate change. To investigate this further, we conducted research on how a pollutant, CO, would be affected when there assumed to be a decrease in human activity caused by the onset of the current COVID-19 pandemic. The shutdown in March gave us the perfect opportunity to do so. The research conducted is important because with the effects of climate change starting to appear such as the temperature rising more than 2 degrees Fahrenheit in the last century, the sea level rising, the ice caps melting; humans need to find a solution to the problem that we have created. All these changes have been caused by pollution, a product of human activity. The research included mainly data analysis of the CO levels. Increasing pollution levels is one of the most significant issues that faces the current generations. Pollutants such as CO, $CO_2$, $SO_2$, $NO_2$, $PM_{2.5}$, $PM_{10}$, and Ozone are the most problematic and dangerous to agriculture, human health, and the environment. Authors claim that they "know a great deal about the direct or primary effects of air pollutants on symptom expression, growth, and yield. This is the part of the iceberg that is obvious and readily seen" (Webb, Taylor, and Tingley 365) (1). It is important to understand that these pollutants only affect air pollution. Water pollution, land pollution, noise pollution, plastic pollution, and soil contamination are the other plagues affecting us. For example, water pollution leads to problems within marine ecosystems and disrupts it (2). Air pollution has the worst effect on human health. These problems range from minor upper respiratory irritation to chronic respiratory and heart disease, lung cancer, acute respiratory infections in children and chronic bronchitis in adults, aggravating pre-existing heart and lung disease, or asthmatic attacks(4). In addition to health problems, air pollution also has devastating effects on the environment (4). Some of them being acute morbidity in various species of trees and soil and forest contamination (4). Mining, deforestation, factories, power plants, airplanes, cars and the burning of fossil fuels are the main contributors to the pollution we observe today (5) (6) (7). All these activities are carried out every single day globally which creates a ton of pollution. Air pollution is a global problem that's only getting worse (8). CO, one of the pollutants is very harmful to humans, mainly due to Carbon Monoxide poisoning. With pollution rising, the CO levels have also gone up in the past decades (9).

Therefore, our hypothesis was that the CO levels would decrease with the decrease in human activity, assumedly driven by the limitations put in place during the beginning of the current pandemic that we are currently enduring. The data analysis conducted brought us one step closer to inferring how a potential decrease in human activity would affect CO levels.





## METHODS

The data used in the experiment was obtained from the EPA's Air Quality public data available on its website. The CO levels were downloaded among other pollutants such as $SO_2$, $NO_2$, Ozone, Pb, $PM_{2.5}$, and $PM_{10}$. The Region for the data was selected to be the state, which included every site available in the state. The number of these sites depended on the state, with more populous states having more sites and less populous states having less sites. After the data was obtained, the datasets were edited to make the analysis more understandable and cleaner. The original datasets included the specific dates such as 1/1/2020 but these dates were changed to the number of days since 2020, so 1/1/2020 would be 1 and 1/31/2020 would be 31. The other thing that was changed was the label for the CO values. Originally, it was Daily Max 8 hour CO Concentration but it was changed to Daily. Max.8.hour.CO.Concentration to avoid errors in R since it's space sensitive. This was done to make the graph easier to read and to analyze the data better. This was done for each site to make sure that the data was still accurate since the dates were an important part of the dataset. After the above steps were done for every state, all the edited datasets were imported to R studio. The additional packages used were: ggplot2, dplyr, plotrix, and ggpubr.

| Test | Value | Analysis |
|---|---|---|
| Cor Test | -0.3569 | Strong Correlation |
| $R^2$ Value | 0.127 | Significant Finding |
| P Value | 2.2e-16 | Strong Correlation |

**Table 1** Test results of the data analysis in R.

The datasets were imported using the read_csv function. This allowed the dataset to be analyzed. Then, the datasets were filtered since the original downloads had unnecessary information such as Site ID, County location, County ID. This could also have been done by simply removing these columns while the datasets were being edited. Since there were only three columns that were necessary, only those ones were selected and further analyzed. The exact code and an explanation can be viewed here: **Code file**.

## RESULTS

The effect of the shutdown on CO levels was examined through analyzing the mean from hotspots across the United states. Data was collected from the Environment Protection Agency's (EPA) Outdoor Air Quality Data website for 15 states and the mean

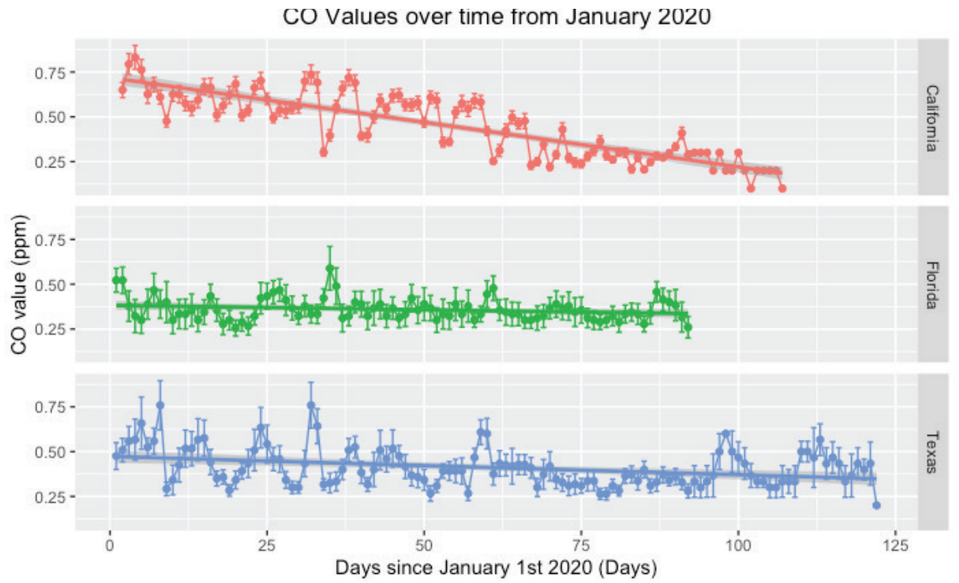



**Figure 1** CO levels in California, Florida, and Texas since January 2020.

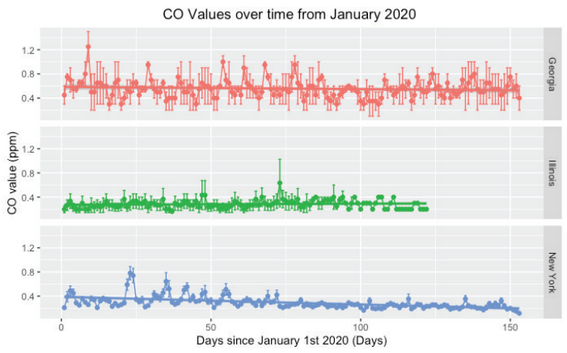



**Figure 2** CO levels in Georgia, Illinois, and New York since January 2020.

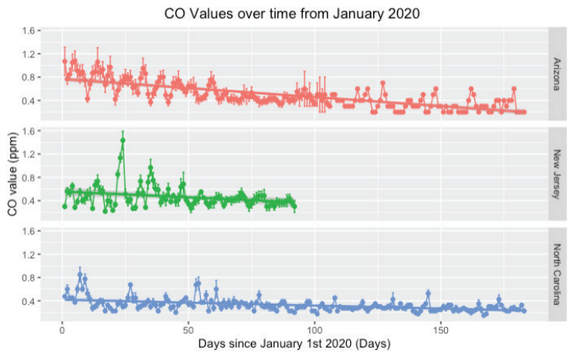



**Figure 3** CO levels in Arizona, New Jersey, and North Carolina since January 2020.

for each state was calculated for the data available. Most of the states showed a decline in CO levels from the beginning of the year. (Figure 1) (Figure 2) (Figure 3) (Figure 4) (Figure 5) For example, in California the shutdown began on March 19th, or the 78th day. In the graph, we see a decline and less fluctuations after the 78th day, meaning that the CO levels dropped and did not fluctuate like they did before the shutdown. (Figure 1) The same is seen for states such as California, Pennsylvania, Arizona, New York, Louisiana, and Tennessee. Data from all states were combined and analyzed to determine if a correlation existed between the CO levels and time since the beginning of the year. A regression analysis was used to determine what proportion of the variation in CO levels was explained by the Dates variable. R Studio's tests were used to analyze the data. The tests used were: A cor test, the R2 value and P value. The cor test gave a result of -0.3569, which was a weak - moderate correlation. (Table 1) A correlation value of -1 is a perfect value since that has a perfect downward slope. In the case of this research, since we were expecting the CO levels to decrease, the results are in line with our hypothesis. If a correlation value is close to 1, then it has a strong positive correlation. The adjusted R2 value was 0.127. (Table 1) This proved the shutdown dates and CO levels had a degree of variation. The P value was 2.2e-16, which indicated that the data had a correlation since the value was very small. (Table 1) More states and more dates would be needed to make a more comprehensive and concrete analysis.

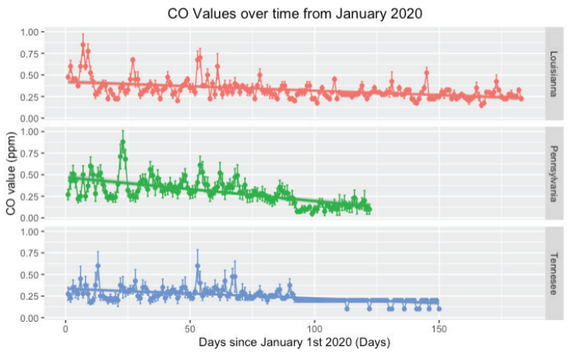



**Figure 4** CO levels in Louisiana, Pennsylvania, and Tennessee since January 2020.

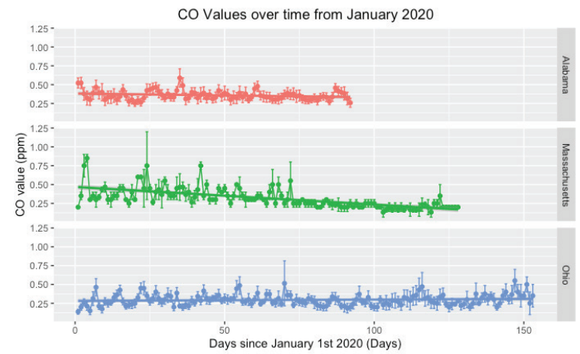



**Figure 5** CO levels in Alabama, Massachusetts, and Ohio since January 2020.

| State | Shutdown Date/Days since January 2020 |
|---|---|
| California | March 19/78 |
| Florida | April 3/93 |
| Texas | April 2/92 |
| Georgia | April 3/93 |
| Illinois | March 21/80 |
| New York | March 22/81 |
| Arizona | March 31/90 |
| New Jersey | March 21/80 |
| North Carolina | March 30/89 |
| Louisiana | March 23/82 |
| Pennsylvania | April 1/91 |
| Tennessee | March 31/90 |
| Alabama | April 4/94 |
| Massachusetts | March 24/83 |
| Ohio | March 23/82 |

**Table 2** Shutdown dates for each state.

## DISCUSSION

The data analysis done showed a decrease in CO levels from the start of the year in January, indicating that the shutdown orders, the earliest one in March, did have a significant effect on the level of CO produced. (Figure 1) (Figure 2) The results from the figures say that human activity largely contributes to pollution. The hypothesis that the CO levels will decrease was supported by most of the 15 states. (Figure 1) (Figure 2) (Figure 3) (Figure 4) (Figure 5) This was confirmed using the cor test, R2 value, and P value since they all showed a correlation between CO levels and the dates.

The cor value was tested on a graph with all 15 states, and gave a result of -0.3569, since the value was close to -1, the correlation was present, but not very significant. Some states saw a higher correlation too, (Table 3) indicating that some states had an almost perfect downward slope since a correlation of -1 is a perfect downward slope. This was mostly because these states had a higher drop in CO levels, primarily because they either had stricter shutdowns or there was less human activity then the other states such as Florida, where the cor value was -0.2166. states such as Florida either did not follow the shutdown orders or had less enforced shutdowns. Looking closely at the data, we see that even though the CO levels didn't drop much in some states, the deviation present before the lockdown was greatly reduced if not absent in Florida, New York, Arizona, New Jersey, North Carolina, Tennessee, and Massachusetts. (Figure 1) (Figure 2) (Figure 3) (Figure 4) (Figure 5) While this was the case in some states, other states such as California, Texas, Arizona, New York, and Pennsylvania experienced significant drops in the CO levels, mostly because these states were more populated and there was a lot more human activity in these states which was suspended. Another interesting thing to notice is the rise of the CO levels in Arizona. Arizona closed on March 31st, around 90 days into 2020. In the graph, we see a decline in CO levels around 75 days. The CO levels stay low until day 100 and around day 120, they start fluctuating again like they were until day 75. Surprisingly, Arizona reopened on May 8th, the 129th day of 2020. This included the opening of casinos, food and drink places, recreation, and personal care. This clearly shows that the CO levels were affected by human activity. The same thing is seen in Ohio, which reopened on May 12, the 132nd day, where we clearly see the CO levels rise. Other states that saw a rise include Texas, Georgia, and North Carolina. This decrease and increase in CO levels is only seen because there was enough data for Arizona, Ohio, Texas, Georgia, and North Carolina to graph this. This would also have been the case for other states had there been enough data. Even with all these tests, the data isn't perfect. There's several factors that might have affected the data, some of them being people not following the shutdown orders, not enough data for each state, exceptions in the shutdown orders, and data from only one source. Since the research was done remotely, it would be difficult to measure independent data for each of the 15 states, which is why the EPA data was chosen. The EPA's data did not have every date available and the same number of sites for all states, instead it either had more sites available for some states and fewer sites for others or fewer dates for some states with more dates for other state, for example, Arizona and New York had around 150 dates while California had around 90. Even with all the correct data, the factors influencing the data might have affected the research. The shutdown orders applied to all public places, offices, schools, and factories. Even with the shutdown orders, there might still be people who don't follow the orders or continue non-essential tasks. In this case, the shutdown order doesn't apply to this group of people as their commute/activity still influences the pollution levels. Finally, there are always exceptions to research, it's no different this time. Even with the shutdown orders, some essential businesses and

| State | Cor Value |
|---|---|
| California | -0.8549 |
| Pennsylvania | -0.6576 |
| Arizona | -0.7767 |
| New York | -0.54416 |
| Louisiana | -0.4983 |
| Tennessee | -0.5769 |

**Table 3** States with the highest cor values.

places need to stay open. These include grocery shops, postal services, agriculture, essential travel, hospitals, and certain factories. For example, when the US had a shortage of ventilators, GM, Tesla, and Ford had their factories working, which means the pandemic did not hinder their operation. The other businesses use a ton of resources and still pollute, making the data not entirely accurate. However, even with all these factors, the research still shows a significant correlation between human activity and CO levels. Other experiments and researches also show a correlation between pollutants and pollution levels and human activity including a study done in one of the most polluted cities in the world. (10) (11) (12) (13) (14). Further research on the effect of human activity on CO levels would further contribute and prove this research. Furthermore, since this research was done in the first few months of the shutdown, it would need to be done as the states start reopening to ensure the results are the same. Since only 15 states in the United States were analyzed, a future study consisting of more locations throughout the world would enhance the study. All these changes would further determine if human activity plays a role in CO levels. Future applications include environmental analysis and climate change studies.

## ACKNOWLEDGEMENTS

I would like to thank my father and my mother for helping me with the review process during the project. I'd also like to thank Spencer Eusden for all his help with the project and for reviewing the manuscript.

## REFERENCES

1. Heck, Walter Webb. et al. Assessment of Crop Loss from Air Pollutants. Elsevier Applied Science, 2013.

2. Guldberg, Ove Hoegh, and John F Bruno. "The Impact of Climate Change on the World's Marine Ecosystems". Science , vol. 328, no. 5985, 18 June 2010. doi:10.1126/science.1189930.

3. Kampa, Marilena, and Eliasz Castanas. "Human Health Effects of Air Pollution". Proceedings of the 4th International Workshop on Biomonitoring of Atmospheric Pollution (With Emphasis on Trace Elements), vol. 151, no. 2, 23 July 2007. <www.sciencedirect.com/science/article/abs/pii/S0269749107002849#!>

4. Smith , William H. "Air Pollution—Effects on the Structure and Function of the Temperate Forest Ecosystem". Environmental Pollution , vol. 6, no. 2, Feb. 1974, pp. 111–129. <www.sciencedirect.com/science/article/abs/pii/0013932774900275>

5. Goudie, A. Human impact: man's role in environmental change. United States: N. p., 1982. Web.

6. Lin, C.-Y. Cynthia, et al. "The Effects of Driving Restrictions on Air Quality: São Paulo, Bogotá, Beijing, and Tianjin." 2011, ageconsearch.umn.edu/record/103381/.

7. Karl, T. R. "Modern Global Climate Change." Science, vol. 302, no. 5651, 2003, pp. 1719–1723., doi:10.1126/science.1090228.

8. Li, Xiangdong, et al. "Air Pollution: a Global Problem Needs Local Fixes." Nature, vol. 570, no. 7762, 25 June 2019, pp. 437–439., doi:10.1038/d41586-019-01960-7.

9. Ernst, Armin, and Joseph D. Zibrak. "Carbon Monoxide Poisoning." The New England Journal of Medicine , 26 Nov. 1998, <www.nejm.org/doi/full/10.1056/NEJM199811263392206>

10. Bao, Rui, and Acheng Zhang. "Does Lockdown Reduce Air Pollution? Evidence from 44 Cities in Northern China." Science of The Total Environment, vol. 731, 2020, p. 139052., doi:10.1016/j.scitotenv.2020.139052.

11. Wang, Qiang, and Min Su. "A Preliminary Assessment of the Impact of COVID-19 on Environment – A Case Study of China." Science of The Total Environment, vol. 728, 22 Apr. 2020, p. 138915., doi:10.1016/j.scitotenv.2020.138915.

12. Myllyvirta, Lauri. "Analysis: Coronavirus Temporarily Reduced China's $CO_2$ Emissions by a Quarter." Carbon Brief, 26 June 2020, www.carbonbrief.org/analysis-coronavirus-has-temporarily-reduced-chinas-co2-emissions-by-a-quarter.

13. Chauhan, Akshansha, and Ramesh P. Singh. "Decline in PM2.5 Concentrations over Major Cities around the World Associated with COVID-19." Environmental Research, vol. 187, 5 May 2020, doi:10.1016/j.envres.2020.109634.

14. Kerimray, Aiymgul, et al. "Assessing Air Quality Changes in Large Cities during COVID-19 Lockdowns: The Impacts of Traffic-Free Urban Conditions in Almaty, Kazakhstan." Science of The Total Environment, vol. 730, 4 May 2020, doi:10.1016/j.scitotenv.2020.139179.



HEADWATERS SCIENCE INSTITUTE
Student Research Journal

# Is there a pattern between genetics and future growth/reduction in the ecological niche of a species within the *Cypripedium* genus? - Insights based on remote genetic data and ecological niche modeling

**Lauren Lee[1] and Daniel Dudek[2]**
[1]Phillips Academy Andover, [2]University of Texas Arlington

## ABSTRACT

**This research seeks to examine the genetic relationships between eight species within the *Cypripedium* genus, using data from genbank, as well as to create ecological niche models for each of these species. The species were selected mostly based on conservation status - most of them are classified by the IUCN red-list as either vulnerable or endangered, with a decreasing population. The ecological niche modeling created a geographic visual for each species' ability to thrive in its indigeonous location in both 2020 and 2050, therefore showing the species propensity to either expand or reduce as the climate in each of the regions is projected to change according to current climate models. These two sets of data were compared in an effort to explore how genetically different the species selected were and whether or not genetically similar species experienced similar fluctuations in ecological niche preference caused by climate change. No such pattern could be discerned.**

## INTRODUCTION

This project was conducted under the context of the Headwaters Science Institute Summer Research program - an educational program for high school students to practice research skills based in the California bay area. The aim of this project was to learn more about how species are related at the genetic level and ecological preferences for a given species through the execution of basic phylogenetic methods and Wallace, a user-friendly ecological niche modeling tool.

Since 1900, an average of three seed-bearing plants have begun to go extinct per year - this not including plants which are extinct naturally or with populations so small that they are not expected to ever recover. The rate at which this is occurring is "500 times higher than would be expected as a result of natural forces alone (1)."

The goal of this experiment is to find whether or not there is an observable pattern between the endangerment, or a trend of decrease in habitability within a species' native region, and that species' genetic makeup. Focus was centered on the *Cypripedium* (lady's slipper) genus for several reasons, but lady's slippers are a plant that I find endearing due to their beauty and family lore. Numerous species of *Cypripedium* are currently either vulnerable or endangered, and they occupy a wide global range. The *Cypripedium* genus belongs to the Orchid plant family, which is known to have higher rates of rarity than others (2-4). The combination of plant rarity and a decreasing population could well suggest that many species will soon become extinct. Although *Cypripedium* is rare, it does play a role in the natural food chain - it is a well known fact that deers commonly consume them. Specifically, *Cypripedium Calceolus* seems to be at a high risk of extinction, and is currently "one of the most intensely studied European plants (8).

There are many factors that could cause any of the eight species of *Cypripedium* that were studied to either increase or decrease in future populations. Of the eight species of Cypripedium selected, six (*C. Kentuckiense, C. Californicum, C. Macranthos, C. Japonicum, C. Parviflorum, and C. Passerinum*) currently have a decreasing population, and four (*C. Kentuckiense, C. Californicum, C.Japonicum, and C. Passerinum*) are either vulnerable or endangered (5). However, when plant species are introduced to a new environment, it can begin to display genetic change from the original species due to the effects of local adaptation, so a plant's genetic difference

from it and it's recent ancestor might explain the growth of the species within a certain region and a reduction within another (6).

Using a variety of software, both a phylogenetic study and ecological niche models were created. Using occurrence and climate data from open resources, ecological niche modeling can create a statistically significant map of suitable habitat for a species within a certain, user selected region. This map can be seen from both a current perspective, and results can be projected into the future - which allows us to predict whether or not a plant's potential ecological preferences will expand or shrink. If their ecological preferences shrink, we can predict that the potential for extinction might increase. Both the phylogenetic tree and ecological niche modeling were compared in an attempt to find patterns between the two.

## METHODS

### Species Selection

Eight *Cypripedium* species were selected for analysis. They include *C. Californicum, C. Kentuckiense, C. Reginae, C. Calceolus, C. Passerinum, C. Parviflora, C. Japonicum, and C. Macranthos*. They were selected in an effort to maximize the wide geographical range in which *Cypripedium* can be found, however only those plants from the Northern Hemisphere could be used as research was limited by the data that was available through a GenBank blast.

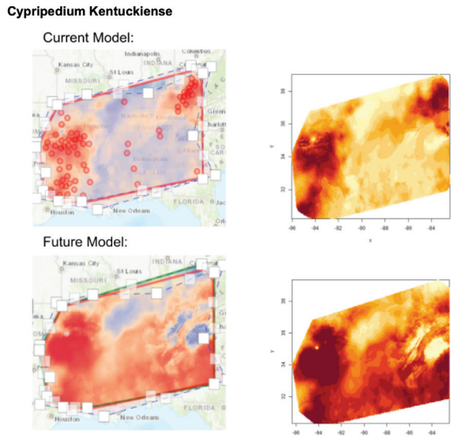



**Figure 2** Cloglog map predictions for ecological niche of C. Kentuckiense, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. C. Kentuckiense demonstrates one of the most clear increases in expected habitability of all species.

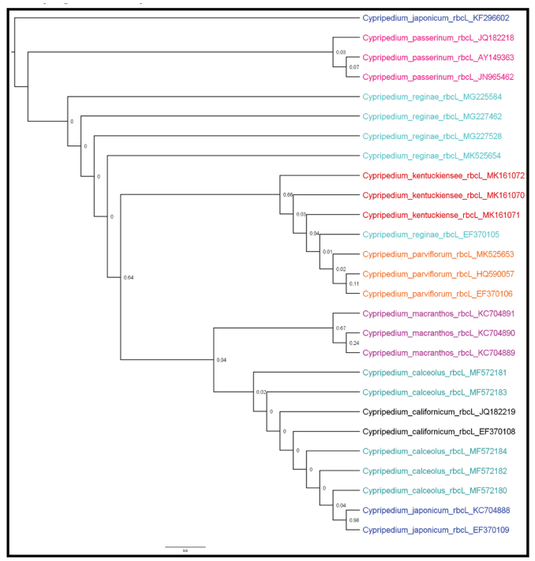



**Figure 1** Above is the resulting maximum likelihood tree. Branches were proportionally transformed for visualization. Sequences were colored according to reported species.

### Phylogenetic Analyses

MEGA-X (Molecular Evolutionary Genetics Analysis) was used to build the phylogenetic tree. By performing a GenBank query, RBCL nucleotide sequence information was obtained for two to five samples representing each species. The sequences were aligned by ClustalW. Parameters for Gap Opening Penalty and Gap Extension Penalty in both Pairwise and Multiple Sequence Alignment were set at 15.00 and 6.66 respectively. Transition weight was set to 0.50, and the Delay Divergent Cutoff was 30%.

The Phylogeny Reconstruction was done by Maximum Likelihood method. To test the phylogeny, Bootstrap Method was used with 500 bootstrap replicates. The substitutions type was set to Nucleotide (versus Amino Acid) using the Tamura - Nei model. Uniform rates were used on all sites. Gaps and missing data were partially deleted, with a site coverage cutoff of 95%. The first, second, third, and noncoding sites were the selected codon sites. The Nearest Neighbor Interchange heuristic method was used, with an automatically created default initial tree. There was no branch swap filter used.

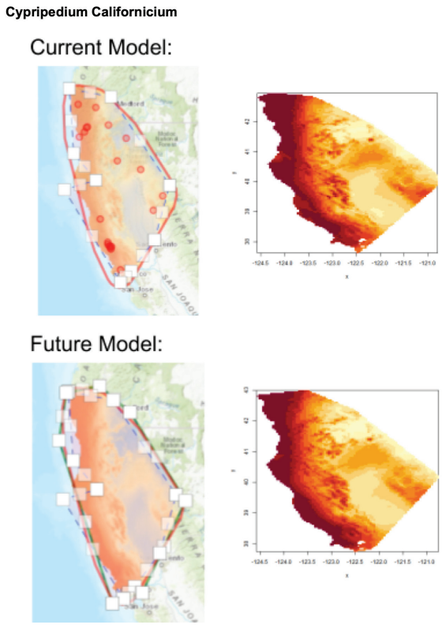



**Figure 3** Cloglog map predictions for ecological niche of C. Californicum, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. Although the change is less dramatic, a slight increase in habitability can still be found on the coastline.

### Ecological Niche Modeling

For each species, computer program Wallace was used to develop an ecological niche model. Wallace is "a modular, R-based platform for reproducible modeling of species niches and distributions (Kass, 2017)." Occurrence data was obtained from GBIF for *C. Calceolus, C. Japonicum, C. Parviflorum, C. Passerinum and C. Macranthos*, and BISON for *C. Kentuckiense and C. Californicum*, as BISON provided more occurrence data. Environmental Data was selected using WorldClim Bioclims with a bioclimatic variable resolution of 2.5 arcminutes. Then, the study region was selected using a minimum convex polygon, and region buffer distance degree of 0.5, sampling ten thousand background points. Occurrence data was spatially partitioned using EnMeval method Block (k=4). The model was then built and evaluated using Maxent, algorithm maxent with feature classes Linear, Quadratic, Hinge and Product, and regularization multipliers between 1.5 and 4.0, with the step value at 1, and clamping enabled. The model results were visualized as a cloglog map prediction with no threshold, and the model chosen was the model with the lowest AICc score, as follows:

*Cypripedium Kentuckiense*: H_2.5
*Cypripedium Californicum*: H_3.5
*Cypripedium Reginae*: LQHP_3.5
*Cypripedium Calceolus*: LQ_3.5
*Cypripedium Macranthos*: LQHP_1.5
*Cypripedium Japonicum*: LQ_2.5
*Cypripedium Parviflorum*: H_2.5
*Cypripedium Passerinum*: LQH_1.5

Finally, a projection model to 2050 was created, using the global circulation model CCSM4, RCP 8.5, and no threshold. Session code can be found in supplementary information.

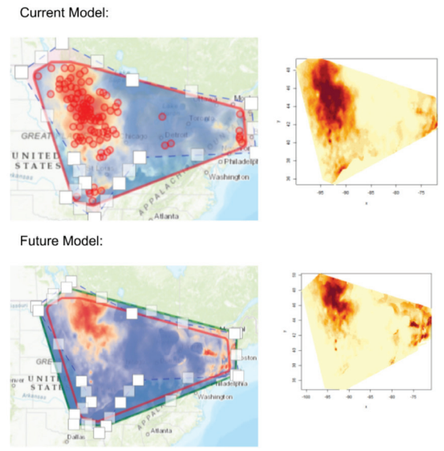



**Figure 4** Cloglog map predictions for ecological niche of C. Reginae, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. There is a clear reduction from the current model to the future model, especially in the upper left corner.

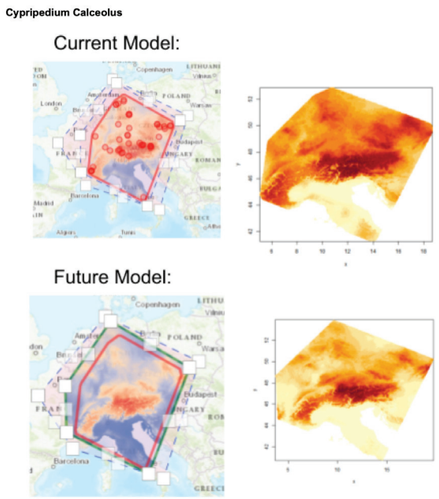



**Figure 5** Cloglog map predictions for ecological niche of C. Calceolus, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image.There is a dramatic reduction from the current model to the future model, comparing color concentration especially towards the tops of the raster images.

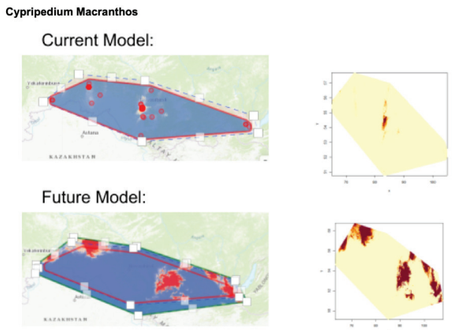



**Figure 6** Cloglog map predictions for ecological niche of C. Macranthos, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. There is a stark increase from the current model to the future model, with several more regions becoming habitable.

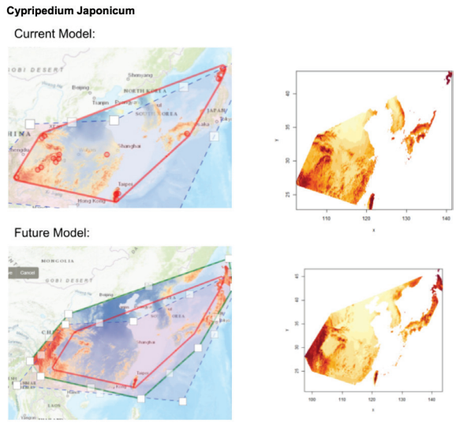



**Figure 7** Cloglog map predictions for ecological niche of C. Japonicum, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. There is a decrease from the current to the future model, which can be observed best in the lower left corner.

## RESULTS

### Ecological Niche Modeling Results:

Session code in R for the following can be found in the supplementary information category. Each species has a current ecological niche model, and future (projected to 2050) ecological niche model visualized in map format. For each species, the model is displayed on the left on a map of the region, and on the right as a raster, without the map. On a continuous scale from 0 to 1, representing probability of occurrence, colors are warmer as the probability of occurrence increases.

## DISCUSSION

The null hypothesis appears to be accurate - the four species that are projected to grow in ecological niche size and the four species that are not are not more or less related on the phylogenetic tree - in essence there is no observable pattern between genetic relatedness and the future viability of a species in the Cypripedium genus. For example, you can see that *C. Japonicum and C. Parviflorum* are least related (yet both are projected to decrease in range by 2050. )

Additionally, it appears that of the eight species, four





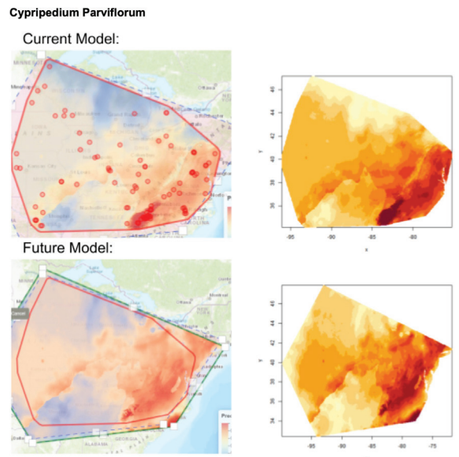



**Figure 8** Cloglog map predictions for ecological niche of C. Parviflorum, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. There is a decrease from the current to the future model, which can be observed best in the lower right corner.

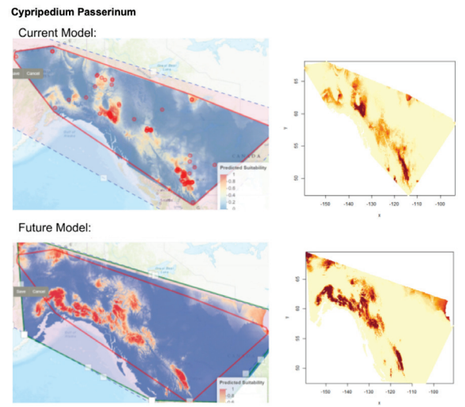



**Figure 9** Cloglog map predictions for ecological niche of C. Passerinum, both current and future. On the left the prediction is superimposed onto a map, on the right is the raw raster image. There is an increase from the current to the future model, which can be observed best in the upper left corner.

(*C. Kentuckiense, C. Californium, C. Macranthos and C. Passerinum*) are projected to expand in ecological niche size in the future, while four (*C. Calceolus, C. Japonicum, C. Reginae, and C. Parviflorum*) are projected to reduce. Ecological niche modeling has been used already to make predictions about the future of *Cypripedium* species, specifically *C. Calceolus* (the results of which corroborate the results of this study) (8). Evidently, the species that are located in Eurasia are projected to shrink while those located in North America are projected to expand, ergo the future expansion of *Cypripedium* species is likely due mostly to environmental or climatic factors. That being said, three of the four North American species share a recent common ancestor, though those three exist in especially different North American Climates (specifically southern America, north-eastern America, and sub-artic regions of Alaska and Canada).

This research is not comprehensive and was a learning experience, but the ecological niche models shed light into how these species ecological niche preferences will change over the course of the coming decades. Future expansions on this project should include: making the research more specific and widespread by including both more species, and using more detailed projection techniques through Wallace, specifically through projection to 2070 in addition to 2050 to see if the same patterns of growth and reduction continue well into the future. Additionally, in order to connect the phylogenetic component to the ecological niche modeling component, modeling only genetically confirmed species would be required.

Within the *Cypripedium* genus only eight of the 58 species were examined. All of the species chosen are indigeonus to the northern hemisphere, and many are native to North America. Even within the few species studied, many will experience drastic population reductions. This change in climate and population is exemplified over a period of only thirty years. Such a reduction will obviously not be limited to this genus, and a rapid decrease in biodiversity can have negative effects on both the wildlife and human life (10).

## ACKNOWLEDGEMENTS

I have immense gratitude for the generosity of Headwaters Science Institute - they provided me with a scholarship for their program, through which I learned everything

that I needed for this project. Specifically, within the program, thank you to Mr. Daniel Dudek - your support and encouragement were invaluable to me - thank you so much for your time.

## REFERENCES

1. Ledford, Heidi. "World's Largest Plant Survey Reveals Alarming Extinction Rate." Nature News, Nature Publishing Group, 10 June 2019, www.nature.com/articles/d41586-019-01810-6.

2. Primack , Richard B. and Hall, Pamela. "Costs of Reproduction in the Pink Lady's Slipper Orchid: A Four-Year Experimental Study." The American Naturalist, The University of Chicago Press, 1 Nov. 1990, www.journals.uchicago.edu/doi/abs/10.1086/285120.

3. Swarts, Nigel D., and Kingsley W. Dixon. "Terrestrial Orchid Conservation in the Age of Extinction." Annals of Botany, U.S. National Library of Medicine, 14 Feb. 2009, pubmed.ncbi.nlm.nih.gov/19218582/.

4. McCormick, Melissa K., and Hans Jacquemyn. "What Constrains the Distribution of Orchid Populations?" New Phytologist Foundation, John Wiley & Sons, Ltd, 12 Dec. 2013, nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.12639.

5. IUCN 2020. The IUCN Red List of Threatened Species. Version 2020-2. https://www.iucnredlist.org. Downloaded on 09 July 2020.

6. Stout, Jane C, et al. "Genetic Diversity and Floral Width Variation in Introduced and Native Populations of a Long-Lived Woody Perennial." AoB PLANTS, Oxford University Press, 19 Dec. 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4323518/.

7. Peterson, Andrew. "Predicting species' geographic distributions based on ecological niche modeling." Condor, August 2001.

8. Kolanowska, Marta, and Anna Jakubska-Busse. "Is the Lady's-Slipper Orchid (Cypripedium Calceolus) Likely to Shortly Become Extinct in Europe?-Insights Based on Ecological Niche Modelling." PLOS ONE, Public Library of Science, 31 Jan. 2020, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0228420.

9. Kass, J. (2017). Wallace. Retrieved August 20, 2020, from https://wallaceecomod.github.io/

10. Unit, B. (2010, March 24). Introduction. Retrieved August 20, 2020, from https://www.cbd.int/climate/intro.shtml



HEADWATERS SCIENCE INSTITUTE
Student Research Journal

# Presence of Vegetation in Relation to Slope in Yosemite Valley, California

**Colin Saltzgaber[1], Julia Kelson[2]**
[1]The Nueva School, [2]The University of Michigan

## ABSTRACT

**Here we examine the relationship between the slope of a terrain and vegetation, measured by NDVI levels. I used satellite data from Yosemite Valley to analyze and test for statistical significance between NDVI levels amongst different slope angles. I used confidence intervals and hypothesis tests to estimate the population mean NDVI levels. I found statistical significance between the 30-45 degree range and the 75-90 degree range, which supports my hypothesis that lower sloped terrain is more supportive of vegetation. However, my findings show that no slope (even as extreme as 85-90 degrees) prohibits the growth of vegetation completely, and that even the steepest slopes analyzed contained plant life.**

## INTRODUCTION

Vegetation growth and presence is controlled by several factors including elevation, parent material, slope, and aspect, all of which change over thousands to millions of years (e.g., 1). High degree slopes are commonly thought to hinder or prohibit vegetation growth, however, even cliffs can actually be supportive of vegetation. A research team analyzed the presence of vegetation among cliffs which returned surprising and unexpected information, revealing how all of the sixty-five cliffs that they sampled across five countries were vegetated and supported woody taxa (2). Similarly, another research team found vegetation at high slopes, however they also discovered that the steep terrain that the plant life was present on disturbed the growth process, resulting in "bent trees and widespread pit-and-mound microtopography" (3). In this work we are treating slope angle as the independent variable, but it is worth considering that vegetation can influence slope stability (and perhaps steepness, by extension); such feedback between slope steepness and vegetation have been long observed by geomorphologists (i.e., 4). I hypothesized that cliffs would be less supportive of vegetation compared to lower sloped areas which therefore led me to analyze this relationship in a local area. More specifically, I chose to investigate the relationship between the slope of the terrain in Yosemite Valley and the vegetation that is present (which I measured using NDVI levels which represents the density of green).

I analyzed data from Yosemite National Park for two main reasons. The first is that it is a place that I've visited many times and truly appreciate, and it's arguably one of the most famous national parks in the world. The second, more scientific, reason, is that Yosemite Valley contains a large range of slopes within a relatively small area. The slopes can range from sheer granite cliffs (such as the famous Half Dome cliff or El Capitan), to flat valleys and meadows (such as Yosemite Valley). These diverse slope angles allow for variance in the data. Also, because Yosemite is primarily formed from granite rock, the weathering profiles vary minimally throughout the study area, therefore serving as a control throughout this experiment. Lastly, since I collect the data at a relatively small area in California, there are no major differences in temperature, moisture, rainfall, etc., except the unavoidable decrease in temperature with elevation. Here I explore how the slope of terrain affects the presence of vegetation, and my hypothesis is that lower sloped terrain will support more vegetation and that the correlation between the two will be negative (i.e. as slope degree increases, vegetation will decrease).

## METHODS

In order to analyze the relationship between the slope of a terrain and its corresponding NDVI (normalized difference vegetation index) levels, I examined two datasets. The slope dataset that I used for Yosemite Valley is LiDAR (light detection and ranging) acquired data which was processed and completed by the National Center for Airborne Laser Mapping (NCALM). This specific dataset was downloaded through Open

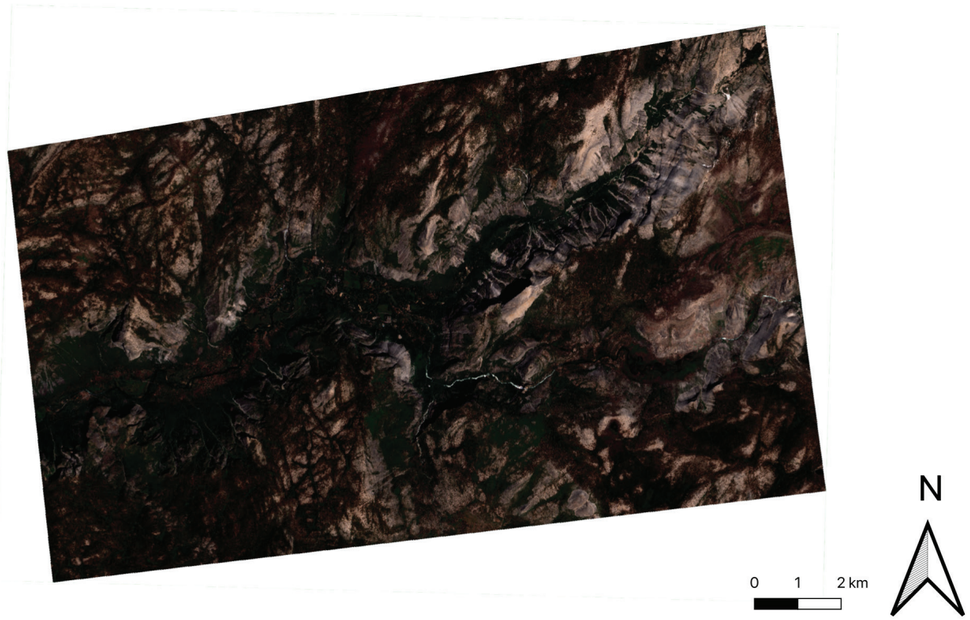



**Figure 1:** Satellite Image of Yosemite Valley. An aerial image of Yosemite Valley captured from satellite imagery. The transects were drawn perpendicular to the general direction of the valley (E-W) which captured many angles of slope.

Topography which contains an elevation model at a 1m resolution which I used to calculate the degree of the slope. The dataset that I used to extract NDVI levels (a good approximation of vegetation, as it measures the density of green on a patch of land) of Yosemite Valley was downloaded through Sentinel Hub (Measuring Vegetation 2000). This dataset is based on satellite data captured at a 10m resolution, containing two specific bands, B8 and B4 (near infrared and visible red respectively), which I used to calculate NDVI levels using the equation, NDVI = index (B8, B4) = (B8-B4)/(B8+B4) (Figure 2)(6). This satellite image was recorded on July 11, 2017. I then overlaid both of the two datasets, projecting the two layers to be the same resolution (10m) as well as in the UTM system, ensuring that for every slope data point, there is a corresponding NDVI data point. This was all done in QGIS, an interactive open-source application designed for geospatial analysis. The area which I examined this relationship from can be seen in Figure 1. Within Yosemite Valley, which is an east-west striking valley surrounded by granite mountains and cliffs on both sides, I created four total transects that spanned and encompassed these two different geographic features that are present. I extracted the corresponding data for each of the four specific

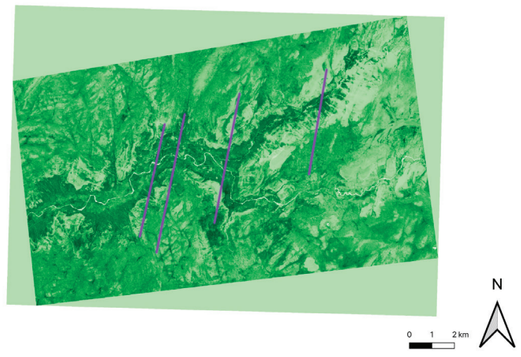



**Figure 2:** NDVI Yosemite Valley. The image is the same in dimension and location as the one featured above. In the valley, darker green shades can be observed (high NDVI values). Low NDVI values are light green. In addition, the transects which were examined are highlighted in purple.





lines which returned two columns of data—one for slope, and the other for NDVI. The coordinates of the four transects that I extracted data from are below, with the format (UTM X, UTM Y): Transect 1, start — (270,404.51, 4,176,919.91), end — (271,549.30, 4,181,989.68), Transect 2, start — (277,978.33, 4,179,687.85), end — (278,747.21, 4,184,198.57), Transect 3, start — (273,728.40, 4,177,506.20), end — (274,797.07, 4,183,205.78), Transect 4, start — (271,049.85, 4,176,165.11), end — (272,387.42, 4,182,351.84).

Finally, to analyze and determine whether slope had an effect on NDVI, I used confidence intervals and hypothesis tests to test for statistical significance between two ranges of slope degrees. I divided the slopes into five categories of 15 degree intervals (e.g. 0-15 degrees, 15-30 degrees, etc.) up until 90. I chose to examine the data in the 30-45 degree range vs. the data in the 75-90 degree range. Choosing these two ranges allowed me to compare the distributions of NDVI as a function of slope.

| Transect + Range | Mean NDVI | Lower Bound | Upper Bound |
|---|---|---|---|
| #1 - 75-90 | 0.219 | 0.181 | 0.257 |
| #1 - 30-45 | 0.489 | 0.464 | 0.515 |
| #1 - 0-15 | 0.402 | | |
| #2 - 75-90 | 0.045 | 0.042 | 0.048 |
| #2 - 30-45 | 0.344 | 0.313 | 0.376 |
| #2 - 0-15 | 0.461 | | |
| #3 - 75-90 | 0.167 | 0.089 | 0.245 |
| #3 - 30-45 | 0.405 | 0.368 | 0.441 |
| #3 - 0-15 | 0.545 | | |
| #4 - 75-90 | 0.220 | 0.162 | 0.278 |
| #4 - 30-45 | 0.438 | 0.415 | 0.461 |
| #4 - 0-15 | 0.287 | | |

**Table 1:** Means and confidence intervals of each transect and the two ranges measured within. This table shows the mean NDVI values and the confidence intervals for the 75-90 range and the 30-45 range for all four transects.

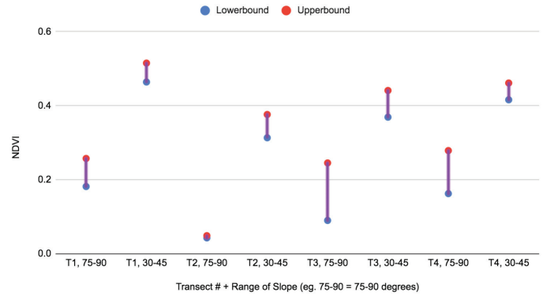



**Figure 3:** Mean and confidence intervals for each transect and range. There are eight confidence intervals total, two for each of the four transects that were analyzed. For each transect, there is an interval for the 30-45 degree range, as well as the 75-90 degree range.

## RESULTS

The hypotheses that I will be using for my statistical analysis and tests are as follows: $H_O$ (null hypothesis) — mean NDVI of 75-90 degrees = mean NDVI of 30-45 degrees, HA (alternative hypothesis) — mean NDVI of 75-90 degrees < mean NDVI of 30-45 degrees, alpha = 0.05 (directly proportional to my confidence intervals which were at a 95% level).

We examined three slope intervals: 0-15, 30-45, 75-90. The interval of 75-90 degrees was chosen as the higher sloped data and the interval of 30-45 degrees was chosen as the lower sloped data. Although I had gathered data that had corresponding NDVI levels for the range of 0-15 degrees, I chose to use the 30-45 degree range for the statistical tests.

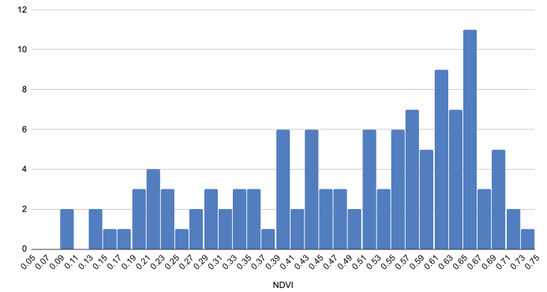



**Figure 4:** Histogram of transect 1 NDVI values in the 30-45 degree range. This histogram shows the distribution of NDVI values that were observed in transect 1 at the specified range. Because of the lower slope, notice the left skewed distribution of the data, implying that it is more common to find higher NDVI values and vegetation at this range.

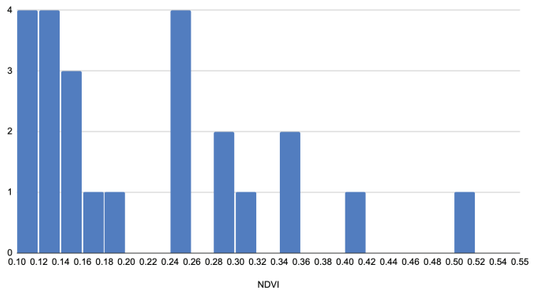



**Figure 5:** Histogram of transect 1 NDVI values in the 75-90 degree range. This histogram also shows the distribution of NDVI values that were observed in transect 1 at the specified range. Because of the higher sloped terrain, notice the right skewed distribution of the data, suggesting that it is less common to find higher values or vegetation at this range.

Confidence intervals and corresponding hypothesis tests which returned p-values are below in Table 1 and Table 2 respectively. Confidence intervals of the means were calculated using the formula $y +/- t^*_{n-1} \bullet SE(y)$, and the hypothesis test for the difference between the means used the formula t-score = $(mean_1 - mean_2) / sqrt((sdev_1 2/n_1)+(sdev_2 2/n_2))$ with a corresponding p-value obtained from a one-tailed test.*

## DISCUSSION

The results support my hypothesis that slope is an influencing factor in NDVI levels, more specifically that higher slopes are less supportive of vegetation than lower slopes (Table 1, 2). I tested the hypothesis using two methods to determine if there was statistical significance between the ranges of slope. Although I had data for all ranges of slopes (0-90 degrees), I chose to not compare the 0-15 degree range (instead I chose the 30-45 degree range) to the 75-90 degree range because after more closely examining the physical landscape that corresponded to the 0-15 degree range, I noticed that there were a lot of roads, gravel patches, houses/hotels, and dry meadows, all of which would ultimately affect the estimation of natural vegetation in those specific areas (e.g. if the grass is dead).

The first statistical test used is to compare the ranges of the confidence intervals. Since confidence intervals and hypothesis tests are directly related, if two confidence intervals fail to overlap (with their respected ranges), there is statistically significant data between the two samples. All four pairs of confidence intervals fail to overlap one another, which means that all of the there is statistically significant data between the NDVI values from high slopes (75-90 degrees) and the NDVI values from lower slopes (30-45 degrees) within a given transect (Table 1, Figure 3). The second test for statistical significance is performing a standard hypothesis test for the difference between means, which were represented through my t-scores for each transect as well as the p-values obtained (Table 2). All of the p-values are lower than the alpha value (0.05), indicating that the samples showed statistically significant differences between the two ranges of slope. This means we can reject the null hypothesis and accept the alternative hypothesis which states that the mean of NDVI for 30-45 degrees is not equal to the mean of NDVI for 75-90 degrees. In addition, I combined all the data from these four transects and got a p-value of 0.0001, which gives the same result as above, in that we can reject the null hypothesis and accept the alternative hypothesis.

The distributions of NDVI in the 30-45 degree interval to the 75-90 degree interval are different from each other. The distribution of NDVI in the lower sloped terrain is left skewed, while the distribution of NDVI in the higher sloped terrain is right skewed. This implies that for the 30-45 degree range, NDVI values lower than the mean are less common, whereas for the 75-90 degree range,

| Transect | Degrees of Freedom | T-Score | P-Value |
|---|---|---|---|
| #1, 30-45 vs. 75-90 | 140 | 10.006 | 0.00001 |
| #2, 30-45 vs. 75-90 | 107 | 15.717 | 0.00001 |
| #3, 30-45 vs. 75-90 | 105 | 5.0396 | 0.00001 |
| #4, 30-45 vs. 75-90 | 166 | 6.1596 | 0.00001 |
| All/Combined Data | 521 | 12.903 | 0.00001 |

**Table 2:** Hypothesis test results comparing the means of the two intervals measured for NDVI. This table shows the results of the hypothesis test for each of the transects, calculated from the degrees of freedom (found by subtracting 1 from the sample size), the corresponding t-score (alpha = 0.05), and the p-value. All/Combined Data is all of the data from the four transects combined.



HEADWATERS SCIENCE INSTITUTE
Student Research Journal

NDVI values higher than the mean are less common (Figure 4, 5). Pearson's coefficient of skewness for the 30-45 degree range and 75-90 range is -0.259 and 0.350 respectively. This portrays the difference in levels of NDVI between these two ranges of data and for how in the higher sloped range there are less observed patches/data containing high vegetation, implying that slope is an influencing factor for vegetation growth.

However, although my analysis suggests that lower sloped terrain is more supportive of vegetation growth compared to higher sloped terrain, it does not suggest that plant life is non-existent in those areas. For all data points in the 75-90 degree range, all NDVI values were still greater than 0, which reveals that cliffs still support vegetation and that there is no point observed where slope hinders or prohibits plant life completely. This supports the findings of Larson et al. (2000). It is also worth noting that because my satellite data was captured at a 10m resolution, the NDVI value reflects an average over 10m, and does not capture individual trees (<10m).

However, because these tests were based off of four transects that were drawn, there are potential sources of error. The first is variability in the data or the chance that my four transects were not the best representatives of the population of Yosemite Valley and all of its data. This may have affected the accuracy of my results to a certain extent, however I believe that this was not a major problem because all of my transects and samples returned that there was statistical significance between the two slope ranges examined.

For potential further research, I believe that a two-way regression, incorporating another influencing factor for NDVI levels (e.g. moisture levels, aspect, etc.) would be very interesting. As mentioned above, although I did my best to control many variables, there is a chance that others may have influenced vegetation levels. This two-way regression, adding another factor, could address that slight variance in my original project (being that the mean values differ slightly amongst transects), and create a more thorough and complete way of modeling what variables impact NDVI. Lastly, this procedure could be applied to evaluate the relationship between these two variables to other landscapes (both with similar or different characteristics such as climates). My research and findings also have implications and can be applied towards other places around the world that have geologic features similar to Yosemite. These include the Rockies, Sierras, and more, all of which are glaciated landscapes, sprung from the interaction of glaciers over a long period of time. This specific type of terrain is found present all across the globe, which contain similar characteristics to those of Yosemite. Investigating how this relationship stands outside of Yosemite Valley would be interesting to explore down the road.

## ACKNOWLEDGEMENTS

I would like to thank Dr. Julia Kelson for being an amazing mentor and supporting me throughout this project. I would also like to thank Spencer Eusden for giving me the opportunity to conduct scientific research and pursue this passion of mine over the summer.

## REFRENCES

1. Marston, Richard A. "Geomorphology and vegetation on hillslopes: interactions, dependencies, and feedback loops." Geomorphology 116.3-4 (2010): 206-217.

2. Larson, D. W., et al. "Evidence for the widespread occurrence of ancient forests on cliffs." Journal of Biogeography 27.2 (2000): 319-331.

3. Pawlik, Łukasz, et al. "Surface processes and interactions with forest vegetation on a steep mudstone slope, Stołowe Mountains, SW Poland." Catena 109 (2013): 203-216.

4. Gilbert, Carl Grove, and Clarence Edward Dutton. Report on the Geology of the Henry Mountains. US Government Printing Office, 1880.

5.

6. Fragoso-Campón, Laura, et al. "Overstory-understory land cover mapping at the watershed scale: accuracy enhancement by multitemporal remote sensing analysis and LiDAR." Environmental Science and Pollution Research 27.1 (2020): 75-88.

# Investigating the Relationship between Root Carbon Pools and Soil Organic Carbon Pools

**Sabrina S. Ahmed[1] and Daniel Dudek Jr.[2]**

[1]BASIS Independent Silicon Valley, [2]University of Texas Arlington

## ABSTRACT

**Two previously conducted studies (4, 7) investigated the effects of root C pools on soil organic C (SOC) pools. The hypothesis was that root C pools would have a direct positive relationship with SOC pools. When the studies were individually analyzed, it was found that there was a direct relationship between root C pools and SOC pools in the study completed by Dietzel et al., but not in the one by McGroddy. This showed that root C pools had a direct relationship with SOC pools in some areas, but it was not proved to be universally applicable.**

## INTRODUCTION

Soil stores a substantial amount of the carbon found in terrestrial ecosystems (1). To better understand the relationship between root carbon and the levels of carbon in the soil, I conducted an experiment to determine if there is a correlation between these two components. With rising levels of carbon dioxide in the atmosphere, finding more permanent solutions is imperative to slow climate change. If there was a way for plants to grow deeper roots, that could act as a more effective way to store carbon in the earth. Roots that grow closer to the soil's surface are quickly degraded due to the microbes that are present at that depth (2, 3). Having deeper roots would allow plants to store carbon for longer periods of time as they have slower degradation rates (1).

As roots grow in the soil, the contact between the soil and root tissues creates a root C pool. A root C pool is carbon found in anything that can be identified as a root. Soil organic C (SOC) pools are the rest of the organic carbon found in the soil (4). However, it is not clear yet whether root C pools have a direct relationship with SOC pools. Plant roots use granules to sense gravity and grow downwards. Signals that are transmitted when a root senses nutrients in a different direction or when they reach a physical barrier determine how deep and how horizontally spread out the root will be (2). If roots are more concentrated towards the surface, their root C pools are also concentrated towards the soil's surface. Studies have shown that it's possible to alter the way roots grow, making them grow vertically deeper rather than horizontally, without modifying the plant's genetic makeup (5). By changing the roots' vertical depths, it could change the depths of the root C pools, and possibly the SOC pools. There could also be other benefits to growing deep roots, such as reducing water stress (6). However, the relationship between root C pools and SOC pools must be further investigated before considering this as a viable solution to curb climate change.

If there is a direct relationship between the root C pools and SOC pools, root depths can be manipulated to store carbon deeper into the earth. Soil is already an effective carbon sink, but by altering roots to grow deeper, it could not only allow plants to store C for a longer amount of time, but it could also shift the SOC pools to deeper in earth's surface, making it a more reliable and long-term solution to climate change. It can be difficult to gain public support for GMOs, so the alternative that recent studies have found can be a more palatable solution to change the way roots grow.

The hypothesis was that root C pools and SOC pools had a direct relationship, allowing the alteration of root depths to be explored as an option to slow climate change. Data from outside studies (4, 7) have shown that there is a direct relationship between root C pools and SOC pools, which opens up the possibility of altering root depths to store more C in the soil and curb climate change.

## METHODS

Two different, previously collected data sets were used in this data analysis. The first one used was collected from the Tapajós National Forest in November of 1999 (7). The second data set collected information about maize, unfertilized prairie, and fertilized prairie from 2008 to 2013 (4). For this





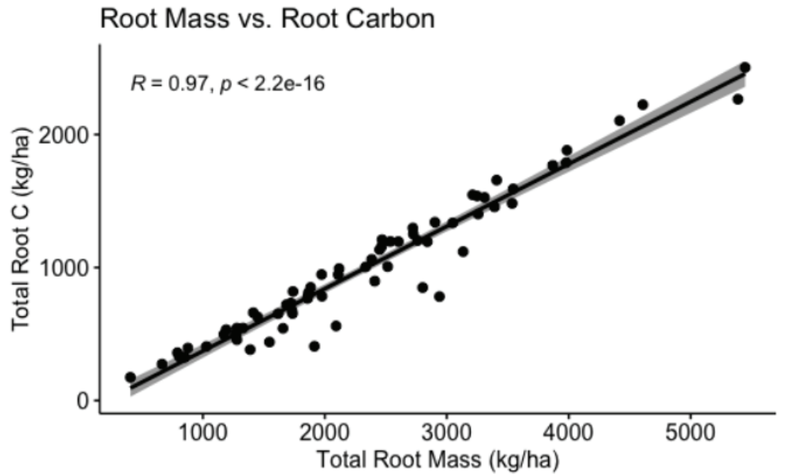



**Figure 1:** There is a strong direct, positive relationship between the total root mass and the total root C. $R^2$ value of 0.941.

experiment, only root C and SOC were examined, although both data sets also looked at other factors.

McGroddy collected the root C and SOC from the Tapajós National Forest through dry combustion. The final data that was reported was a mean of three samples.

Dietzel et al. set up experimental plots which were replicated four times. Three crops were studied in this experiment: maize, unfertilized prairie, and fertilized prairie. Each year, 60-100 grams of pure soil was removed and air-dried. Roots were dried and ground to prepare for combustion analysis, which determined both C and N concentrations.

When organizing and sorting these data sets, we looked only at SOC and root C, but both sets looked at several other variables. We also deleted any rows that did not have a value or that had any extreme outliers. We used R to conduct two separate analyses. Using the ggpubr and ggplot2 packages, We created graphs using the ggscatter function. The summary function was also used to obtain several statistical characteristics, such as the $R^2$ value, p value, and standard deviation.

## RESULTS

The relationship between root C pools and SOC pools was studied through data sets that measured the C content in roots and the soil surrounding them. Two data sets were used; one was collected by McGroddy in the Tapajós National Forest and the other was collected by Dietzel et al. from experimental plots in Iowa. The data sets were analyzed in R. The data collected from the Tapajós National Forest was examined first. It was found that there is a strong positive direct relationship between the root mass and root C (Figure 1). However, there was no strong correlation found between root C and soil C from this data set (Figure 2). With a $R^2$ value of 0.0361 and p value of 0.11, there was no statistically significant relationship between the two factors.

The study by Dietzel et al. yielded different results. The maize and unfertilized prairie both showed a positive direct relationship between the root C pools and the SOC pools, with $R^2$ values of 0.303 and 0.348 respectively (Figure 3). Although the relationship was not as strong as the one between root mass and root C, it was still statistically significant. The correlation between the fertilized prairie root C pool and SOC pool was the strongest (Figure 4). With an $R^2$ value of 0.56, this crop showed a statistically significant correlation.

## DISCUSSION

The hypothesis that root C pools would have a direct relationship with SOC pools was only supported by the study conducted by Dietzel et al. The study conducted by McGroddy confirmed the direct relationship between root mass and root C, but there were no statistically significant results that showed

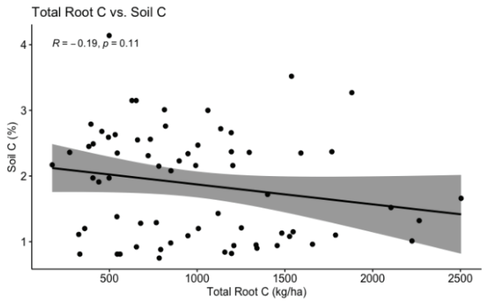


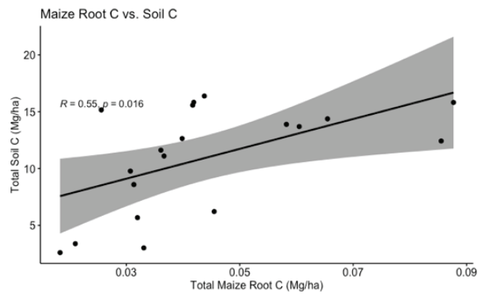

**Figure 2:** There is no strong correlation between the total root C and the soil C % from the data gathered in the Tapajos National Forest, with a $R^2$ value of 0.0361 and p value of 0.11.

a correlation between root C pools and SOC pools. The difference between the two studies could be due to several different reasons. The data collected by McGroddy was from the Tapajós National Forest in the state of Para, Brazil; meanwhile the data collected by Dietzel et al. was collected in the state of Iowa, U.S. The data sets were collected in two geographically different regions. In addition, they measured different plants. McGroddy measured data from the trees and litter of the forest, whereas Dietzel et al. measured the crops planted in the controlled environment. The two data sets also took samples from different depths. Data collected by McGroddy was collected at 10 cm whereas data collected by Deitzel et al. were collected from varying depths.

The methodology conducted in this data analysis could have also led to error. I only studied root C and SOC, but there are several other factors that affect SOC that I didn't take into account, such as nitrogen levels.

Although the data collected from the Tapajos National Forest did not support the hypothesis, the other data set showed a correlation between root C pools and SOC pools. Additional research will have to be conducted to see if that relationship is universally applicable, but this may now be explored in the context of climate change. Research could be conducted to study the effects of altering plants to grow deep roots. Root C pools and SOC pools of the altered plant could be measured at varying depths and compared to the naturally occurring version. Degradation rates of the two versions of the plant could also be compared. Additionally, other variables could be studied to gain a better overall understanding of deep root systems, such as drought resilience or nitrogen levels. Altering the plants to grow deep roots could be achieved in different ways; one could genetically modify genes involving root depth mechanisms or introduce hormone-blocking chemicals to make plants grow deeper roots (2, 3). By conducting these experiments, we will gain necessary information about deep root systems that could aid in the battle against climate change.

This analysis was conducted to test if there was a correlation between root C pools and SOC pools. If there proved to be a correlation, it would be possible to conduct further experiments regarding the effects of deep roots on carbon sequestration in the soil. The hypothesis that root C pools and SOC pools would have a direct relationship was supported by the study conducted by Dietzel et al. in Iowa; however, it was not proved to be a universally observed relationship as the studies conducted by McGroddy in the Tapajos National Forest did not yield



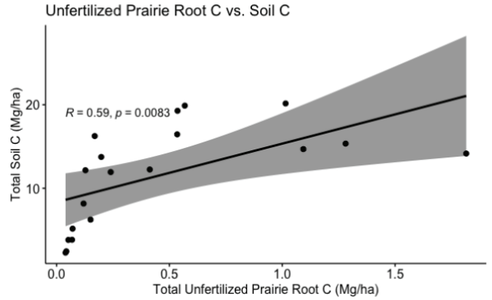




**Figure 3:** There is a direct relationship for both the maize and unfertilized prairie root C pools and SOC pools. The maize has a $R^2$ value of 0.303 and p value of 0.016, and the unfertilized prairie has a $R^2$ value of 0.348 and p value of 0.0083.





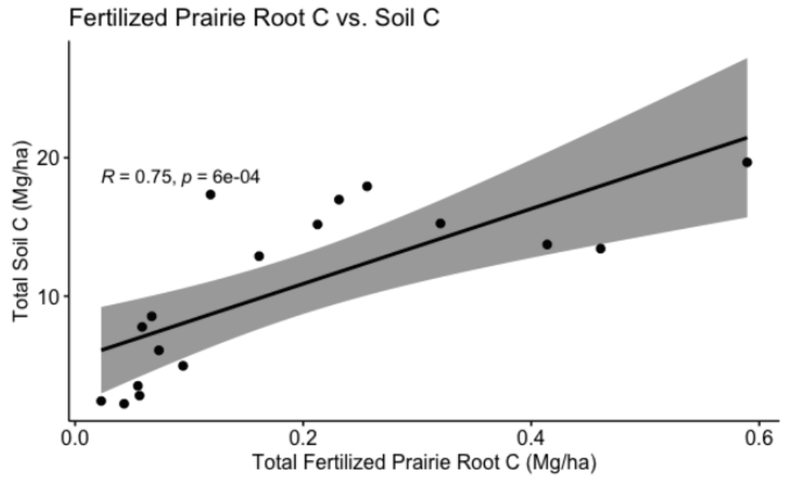



**Figure 4:** There is a stronger direct relationship between the fertilized prairie root C pool and the soil organic C pool, with a $R^2$ value of 0.56.

any statistically significant results that supported the hypothesis.

There were several factors that could have affected the results. For this experiment, only root C and SOC were analyzed, and other variables such as nitrogen levels were not taken into account. In addition, the data was recorded under different conditions. The data collected by Dietzel et al. was from a more controlled environment compared to the data collected by McGroddy.

| Relationship | R | Adjusted $R^2$ | p-value |
|---|---|---|---|
| Root Mass vs. Root C | 0.97 | 0.9379 | < 2.2e-16 |
| Total Root C vs. Soil C | -0.19 | 0.02326 | 0.1119 |
| Maize Root C vs. Soil C | 0.55 | 0.256 | 0.01576 |
| Unfertilized Prairie Root C vs. Soil C | 0.59 | 0.3051 | 0.008336 |
| Fertilized Prairie Root C vs. Soil C | 0.75 | 0.5254 | 0.0005999 |

**Table 1:** Above is a table of the statistical values for each relationship that was studied.

## REFERENCES

1. Ontl, Todd A, and Lisa A Schulte. "Soil Carbon Storage." Nature News, Nature Publishing Group, 2012, www.nature.com/scitable/knowledge/library/soil-carbon-storage-84223790/.

2. McCoy, Berly. "By Making Plant Roots Grow Deeper, These Geneticists Hope to Curb Climate Change." PBS, Public Broadcasting Service, 17 July 2019, www.pbs.org/newshour/science/by-making-plant-roots-grow-deeper-these-geneticists-hope-to-curb-climate-change.

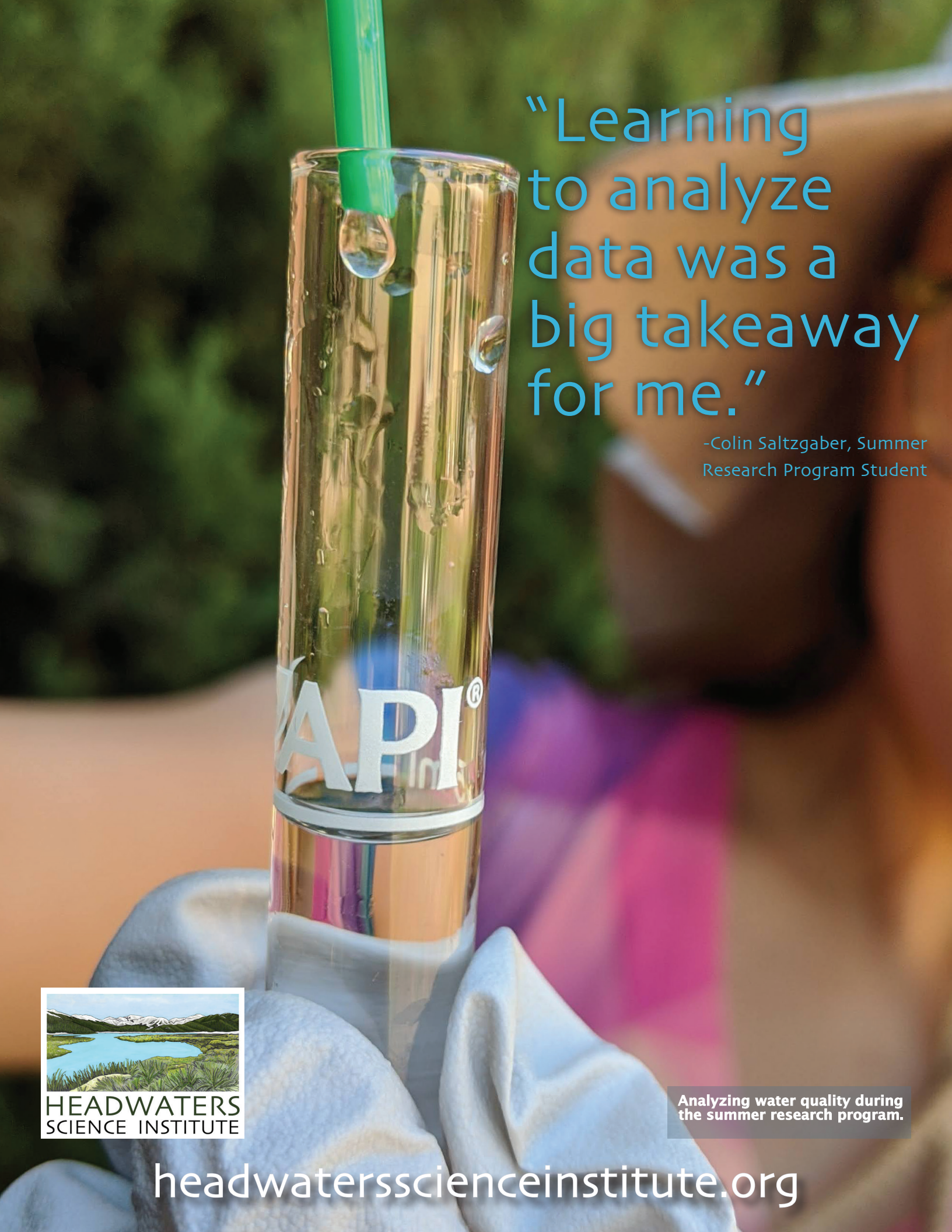
3. Ogura, Takehiko, et al. "Root System Depth in Arabidopsis Is Shaped by EXOCYST70A3 via the Dynamic Modulation of Auxin Transport." Cell, U.S. National Library of Medicine, 2019, pubmed.ncbi.nlm.nih.gov/31299202/.

4. Dietzel, Ranae & Liebman, Matt & Archontoulis, Sotirios. (2017). A deeper look at the relationship between root carbon pools and the vertical distribution of the soil carbon pool. SOIL. 3. 139-152. 10.5194/soil-3-139-2017.

5. Pierik, Ronald, and Christa Testerink. "The Art of Being Flexible: How to Escape from Shade, Salt, and Drought." Plant Physiology, American Society of Plant Biologists, 1 Sept. 2014, www.plantphysiol.org/content/166/1/5.

6. Gewin, Virginia. "An underground revolution: plant breeders are turning their attention to roots to increase yields without causing environmental damage. Virginia Gewin unearths some promising subterranean strategies." Nature, vol. 466, no. 7306, 2010, p. 552+. Gale Academic OneFile, Accessed 28 July 2020.

7. McGroddy, M.E. 2012. LBA-ECO TG-07 Forest Soil P, C, and N Pools, km 83 Site, Tapajós National Forest. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1085.


"Learning to analyze data was a big takeaway for me."
-Colin Saltzgaber, Summer Research Program Student
API®
HEADWATERS SCIENCE INSTITUTE
Analyzing water quality during the summer research program.
headwatersscienceinstitute.org