Executive summary: Soil chemistry is a critical part of ecosystem health and characterization. The interactions between soil chemistry, foliar traits, and climatic data are convoluted and cyclical. Moreover, these features describe and differentiate ecosystems, which form the biomes that allow us to make larger predictions by averaging over similar territories. For my project, I aimed to explore the relationship between soil chemistry and foliar traits at the NEON field sites across the United States which describe 20 unique ecoclimatic zones. For the first section, I replicated the process done by Santos et. al. (2023) to reproduce the linear relationship between soil pH and foliar manganese concentration. However, when I went to examine the target relationship for study, the effect of soil pH and phosphorus concentration on photosynthetic traits (chlorophyll and carotenoid concentration), there was no linear relationship. However, due to the limited data here as compared to the Maire et. al. (2015) article, the null hypothesis can't be fully confirmed. To augment the project, I tested different machine learning classification projects to see if the soil chemistry data could effectively characterize the ecoclimatic zones, which are defined by landforms, vegetation, and ecosystem dynamics. A random forest model successfully classifies the test data with 94% accuracy, suggesting that soil chemistry data alone is enough to train and predict ecoclimatic domain.

Introduction: Plant-soil relationships are complex and defy simple description (Yi 2023, Aitkenhead 2012, Albrechtova 2008, Cornelissen 2010, Maire 2015, Santos 2023). The rudimentary version we learn of their relationship is that the soil has nutrients that the roots absorb and the plant then transports up to the leaf. However, in reality, this process is mediated by climatic variables such as precipitation and temperature, which can affect the solubility of nutrients among other factors, the symbiotic relationships between roots and soil fungi, geochemical soil factors such as soil pH, species-specific traits, and the decomposition of plant matter which returns nutrients to the soil (Yi 2023, Aitkenhead 2012, Albrechtova 2008, Cornelissen 2010, Maire 2015, Santos 2023). The cyclical nature of the process, as well as the large number of variables involved, makes these relationships difficult to understand and to categorize, especially at higher levels than one plant or one site.

Contrary to expectation, the concentration of a nutrient in the soil is not a sufficient indicator of the concentration of the nutrient in the leaf. The presence of nutrients in the soil is mediated by

other factors that promote absorption of the nutrient that can be related to other soil geochemical properties, climatic conditions, or species-specific properties. For example, Cornelissen et. al. (2010) demonstrated that the pH of a leaf is species-dependent rather than soil dependent. Santos et. al. (2023) found that soil pH is a more effective predictor of manganese concentration in the canopy than the concentration of manganese in the soil. Even between the soil and the roots, the level of soil organic carbon (SOC) is both stabilized and destabilized by the roots in a process that is mediated by both geochemical processes in the soil and microbial turnover (Dijkstra 2020).

Understanding which geochemical soil properties are most effective at predicting leaf traits is important for 1) identifying potential mechanisms and limitations for how plants absorb nutrients, 2) identifying which local factors have the greatest effects on how vegetation has adapted to the region, and 3) allowing models to better account for soil variables as well as climatic ones.

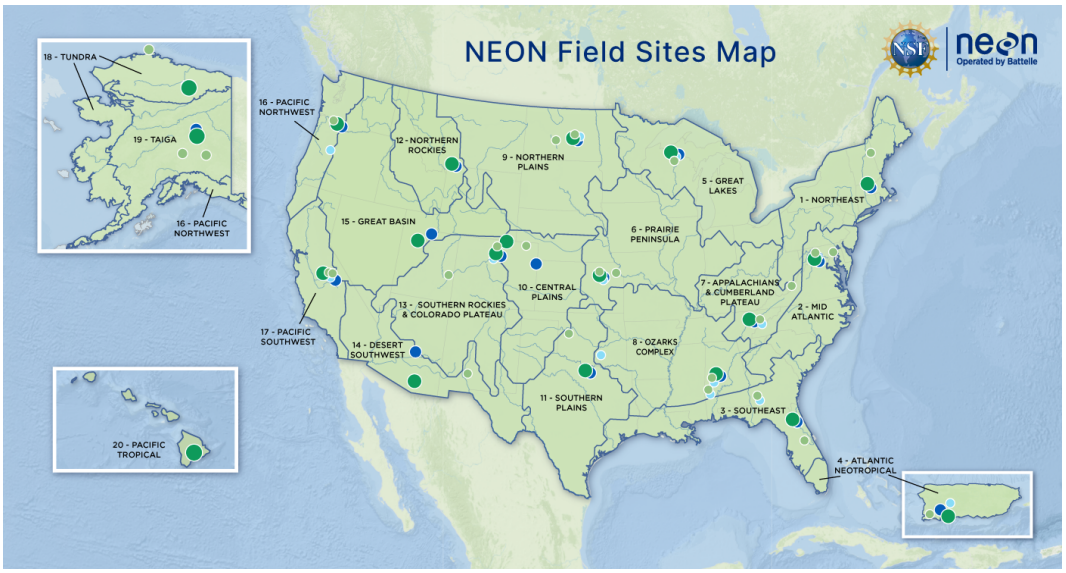

Figure 1: NEON field sites and ecoclimatic domains ("About Field Sites and Domains")

The relationship between soil properties and ecoclimatic zones is also complex and cyclical in nature. Ecoclimatic zones/ecozones/ecoregions are areas of similar ecosystems and environmental resources and are an attempt to group broader regions. The NEON dataset contains 20 ecoclimatic zones across the United States, ranging from the tundra in Alaska to Pacific tropical in Hawaii (Figure 1). As soil chemistry plays a defining role in available land and vegetation for an ecosystem, it plays an important role in shaping each ecoclimatic zone. With climate change, many ecoclimatic zones are experiencing large changes in climate, which can lead to

changes in terrain and potentially soil chemistry. Examining how different environments are able to modulate these changes and how they are changed is of particular importance.

Related Work: The main paper motivating my research topic is "Plant-Soil Relationships Influence Observed Trends Between Manganese and Carbon Across Biomes." Santos et. al. (2023) used data from the NEON database to study the relationship between foliar manganese (Mn), soil carbon (C), organic matter decomposition, and soil pH. The relationship between Mn and C in soils and soil surfaces is the result of complex geochemical processes between plants and soil, and the authors used soil physical and chemical properties and plant foliar traits to study this relationship. They calculated the stock of carbon and nitrogen at each site using data from the NEON sites for each level of soil.

The carbon stock was either averaged or summed for the data analysis. Foliar data was used directly from the sites. The R packages *Hmisc* and *ppcor* were used to get the Spearman rank correlation coefficients between C and N stocks and climatic variables and to examine how their relationship with Mn changed after accounting for those variables. Pearson correlation coefficients were used to evaluate linear relationships among select soil properties among the layers of soil at all of the sites. They were also used to reevaluate the relationships in foliar data. Principal component analysis was used to describe the correlations. Foliar data was compared to soil properties and climatic data to get the relationship between Mn in the soil and Mn in the litter.

Results/Evaluation: The results largely focused on 1) the relationship between Mn concentrations and C and N stocks in soils; 2) the extractable Mn concentrations and pH; 3) the relationship between foliar properties and soil chemistry; and 4) the relationship between foliar Mn and lignin. For the first point, it was found that Mn has an inverse correlation with C and N stocks, which corroborates with existing literature. Secondly, Mn concentrations peaked at around pH 4.5 - 5. Interestingly, foliar Mn is more strongly correlated with soil pH than soil Mn. Lastly, foliar Mn and lignin are highly correlated. Foliar Mn was largely log-transformed for the results, and p and r values were used to examine the statistical significance and strength of the model. These relationships were

also evaluated in the context of surrounding literature and trends were noted across ecosystem types such as evergreen forests or grasslands.

In "Global effects of soil and climate on leaf photosynthetic traits and rates," Maire et. al. (2015) examine how soil pH, available phosphorus, and climatic moisture index affect leaf traits at a global scale. Working at this scale is challenging because of the influence of climatic variables on both soil geochemistry and photosynthetic activity, as well as the large differences in how soil fertility is defined based on the type of ecosystem, where plants have adapted to soil conditions that have very different limiting factors for plant growth. For example, the conditions that limit a desert plant's growth are generally different from the ones that limit plant growth in a rainforest. The researchers use SoilGrids, which is a reliable global dataset that is compatible with global datasets for other ecological trends, such as climate and plant variables. They wanted to examine how soil and climate variables affect plant growth and determine the relationship between statistically relevant predictors and the minimum set of predictors needed to model the most relevant relationships.

Trait data was acquired from the 'Glopnet' database and was supplemented by other sources to create a database from 288 sites and 1,509 species from 165 families. Climatic data was taken from the same source as the foliar data when possible; if not available, it was taken from global datasets. Bioclimatic data was calculated using the climatic data. Soil data was selected for long-term characteristics (ex. soil texture and structure) as opposed to traits that vary within each season (ex. nitrogen mineralization rate). Soil phosphorus was added from other datasets and standardized.

All plant traits were log transformed, and all environmental variables were subjected to the Yeo-Johnson transformation. The relationship between soil pH and soil fertility markers was modeled with a quadratic regression at first, and linear models were used whenever the quadratic term was not significant. Site and species effects were also included, and the models were fitted using the 'lme4' package in R. Each leaf trait was examined using a stepwise mixed regression model to pick the four most important variables (among climatic and soil variables) based on minimizing Akaike information criterion, using the 'MuMIn' R package as well as 'lme4.' They then used variation partitioning and Venn diagrams to partition the variation in each leaf trait into components based only on soil

variables, only on climatic variables, or a joint matrix containing both. Multiple mixed regressions were used as well. The 'vegan' package in R was then used to see how much of the variation in leaf traits could be understood by the matrices of important climatic and soil matrices. Path analysis was done using the 'lavaan' package in R to see how much variation in species is driven by the important climatic and soil traits.

Results: The authors found that soil pH and soil fertility trended as expected based on knowledge of the involved dynamics and the principal dimensions are 1) a carbon and nitrogen component (these are interchangeable) and 2) a phosphorous component. The relationships between leaf traits and soil variables was quantified by examining the relationship between leaf traits and soil traits along the fertility dimensions, and these relationships were evaluated using r squared values. Stepwise multiple regressions were then done and judged by F-values, and the results indicated that soil variables were stronger predictors than climate variables. Variation partitioning was used to compare the effect of climate vs. soil vs. joint effects, with soils emerging as a stronger indicator than climate for leaf trait variation. Redundancy analysis was used to predict the composition and structure of leaf photosynthesis rates from important climate and soil variables: soil pH, available phosphorus, and the climatic moisture index. Path analysis concluded that there were two main dimensions of environmental variation relevant to understanding leaf variation: 1) a soil pH-aridity dimension and 2) the soil-available P dimension.

In "Global soil-climate-biome diagram: linking surface soil properties to climate and biota" by Zhao et. al. (2019) the authors begin to quantify the relationships between soil properties, climatic data, and biome classifications. They combined multiple soil datasets from different databases, containing more than 28,000 soil profiles overall. Due to the high diversity in sources, they looked at eight soil properties: four chemical and four physical traits. Additionally, they had to estimate certain properties, such as bulk density, for large portions of the data (ex. 58%) using other data like soil organic carbon (SOC). The authors used random-forest models to map the specific regions, and then they used that model to predict the spatial distributions of the properties. 10-fold cross validation was

used for averaging, and, once the soil data was mapped, it was averaged for over climatic regions to quantify the linkage between the soil properties and the climatic ones.

The soil map agreed with other known mappings and fit into conventional knowledge about different biomes. For example, arid regions have higher pH than forested ones. The surface soil properties were then placed on the Whittaker climate-biome diagram, creating a soil-climate-biome diagram. The results indicated that climatic variables, such as precipitation, are extremely important in predicting soil properties like pH. Data at the global scale allows for use in larger climate models and helps us to understand the interplay between climate, soil, and biome, which is particularly important in the face of climate change.

Approach: My original approach was to use the methods of the Santos (2023) paper and the reliable NEON datasets to examine the trends in the Maire (2015) paper in sites across the United States, teasing out the nature of the predictive power of soil chemistry on leaf photosynthetic properties. However, I ran into some unexpected roadblocks, and I added a second section of work where I compared different classification models. I chose common models that were covered in depth in class to increase the ease of data analysis.

Results (Section I): For data preprocessing, I followed the NEON instructions to download the rpy2 package to import both general R packages and utilities as well as NEON specific ones. Using these tools, I was able to unzip the data and stack the data for all of the sites into a single file for each dataset. The datasets that I used were the initial soil chemistry dataset ('spc_biogeochem'), megapit soil chemistry dataset ('mgp_perbiogeosample'), foliar chemical dataset ('cfc_elements'), and the foliar chlorophyll dataset ('cfc_chlorophyll'). With each dataset, I removed the non-numeric columns. Additionally, for the initial soil dataset, I created three dataframes: one with all soil horizon data, one with the O horizon data, and one with the A horizon data.

The O horizon is the top layer of soil consisting largely of organic matter. The A horizon is the layer directly below the O horizon and is the first layer of mineral soil. It is the mineral layer with the highest accumulation of plant matter and may be referred to as topsoil. As the top layers of soil, these two contain much of the biological activity that occurs in soil and would be more relevant to plant

activity, which is why the Santos paper began by working with these subsets. Notably, only 24 of the 47 terrestrial NEON sites contained an O horizon and only 44 contained an A horizon.

First, I worked with the O and A horizon data from the initial soil dataset set to replicate the results from the Santos paper to confirm I had selected the proper soil sub horizons. Since their data was site averaged, I site averaged the data and compared the manganese data to the site averages listed in the github repository associated with the paper and the SI.

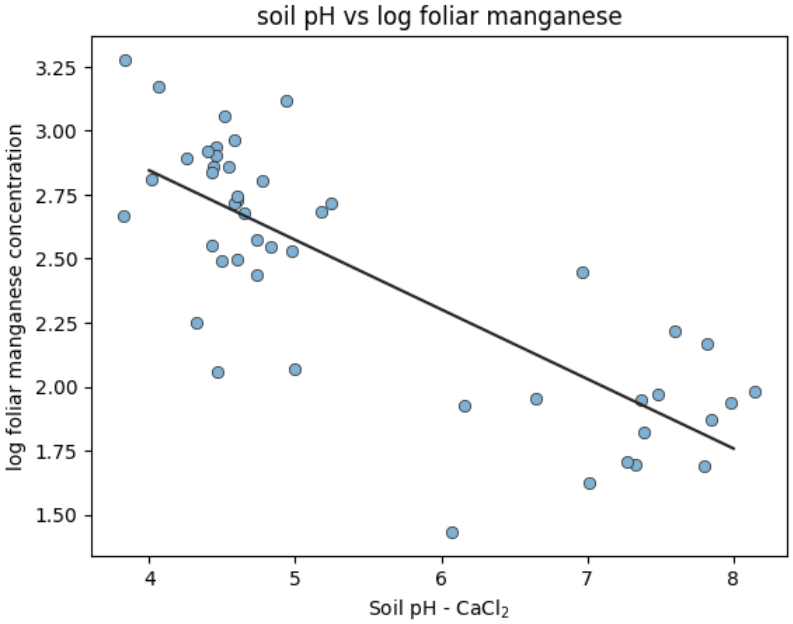

Figure 2: Soil pH vs log foliar Mn concentration, with a linear regression fit. The soil data is from the mega pit dataset

Once I was able to assess that the correct horizons were selected and the averages matched those of the Santos paper, I made sure that there was still a linear relationship between the pH of the soil and the foliar manganese concentration for the megapit soil dataframe. I was able to fit a linear regression to this data with an R-squared value of 0.628 and 0-valued P values, confirming the linear relationship described in the Santos paper (Figure 2). After replicating this result, I wanted to begin applying the Santos methodology to the Maire relationships.

Firstly, I wanted to confirm the Maire paper's assertion that the soil pH and the soil phosphorus were independent and lacked collinearity. I created graphs of the soil pH versus the soil phosphorus concentration for both the O and A soil horizons for the different measures of soil phosphorus: pSatx, which is the phosphate concentration in the < 2mm fraction; pOxalate, which is the soil phosphorus measured by the ammonium oxalate extraction method; Bray1PExtractable, the extractable phosphorus measured by the Bray 1 solution; OlsenPExtractable, the extractable phosphorus measured by the Olsen solution; and MehlichIIITotP, the extractable phosphorus

measured by the Mehlich III solution. For all phosphorus and phosphate measurements, the graphs appear largely diffuse and lack a linear relationship. For the subsequent analysis and relationship exploration, I went with the MehlichIIITotP, as the Mehlich III phosphorus test tends to work for a larger pH range than the other tests (Penn 2018).

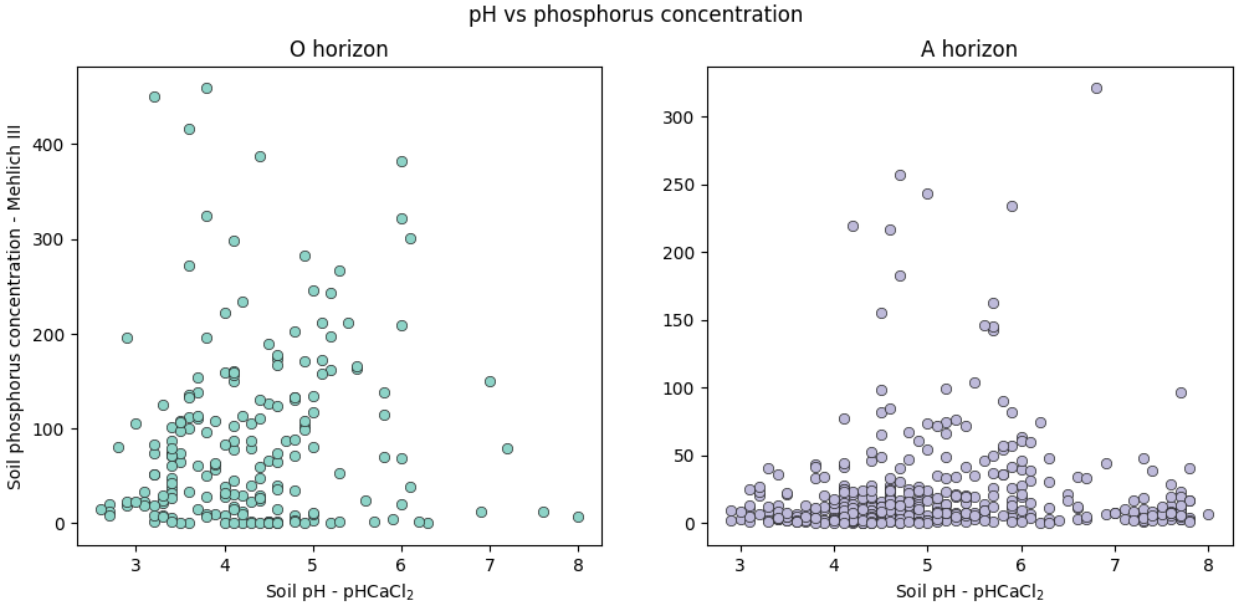

Figure 3: Soil pH vs soil phosphorus concentration, as measured by the Mehlich III extraction, for the initial soil distribution. The left panel shows the data for the O horizon, while the right horizon shows the data for the A horizon

For both the A and O horizons, a linear regression on the Mehlich III phosphorus measurement as a function of the pH had statistically meaningless results, with R-squared values of 0.014 for the O horizons and 0.005 for the A horizons. All variables had high P values in both models. Combining this with a visual inspection of the graphs confirms that these two are not collinear and are independent variables, as described in the Maire paper.

Next, I needed to link the soil chemical data with the foliar chemical data and the foliar chlorophyll data. To do that, I replaced all NaNs with 0 and dropped any columns that contained only "0" entries. Then, I averaged the data across each site. I did this for both the megapit soil data and the initial soil data, although I only split the initial soil data by horizon, with the O horizon and A horizon each getting separate dataframes. Since not all sites had soil from all horizons, the O horizon data only had 24 rows once site averaging was complete, while the A horizon data had 44 rows. The megapit data

had 47 rows, one for each terrestrial site. I examined the relationship between chlorophyll A, chlorophyll B, and carotenoids with soil phosphorus and pH for the initial soil O horizon data. All of these results were statistically insignificant, with R-squared values of 0.054, 0.060, 0.029, respectively. Moreover, all variables had high P values. I tried using the megapit data with the chlorophyll data, and the results also showed a lack of a linear relationship. I tried adding in mixed terms, but all the variables still had high P values and low R-squared values.

Since there are so many more features than rows, I was limited to simpler methods like linear regressions. Thus, I am unable to conclude that soil phosphorus concentration and pH are not linked to plant photosynthetic traits and hold no predictive power. The relationship may be nonlinear or simply not apparent in too small of a dataset. In order to link the values from different datasets, it was necessary to site average, as each dataset has a different number of rows and there is nothing linking one particular foliar sample to one particular soil sample other than the site ID. This reduced the order of magnitude of available data from thousands to 24 - 47 depending on the specific dataset. Since the number of features was in the 50s or more, this combined dataset proved too limited to provide more insight on the relationship between soil phosphorus, soil pH, and photosynthetic data. The only conclusion I feel comfortable making is that, for the initial soil chemical data at the NEON sites, there is not a linear relationship between the target soil chemical data (pH and phosphorus concentration) and the photosynthetic data (the chlorophyll and carotenoid measurements).

With no sense of how to get around the site averaging to augment the dataset and a desire to maintain the consistency of the dataset, I decided to explore relationships among the initial soil data to have a more complete project and get better insight into the data. In particular, I wanted to see how well the 20 ecoclimatic domains are characterized by the initial soil data.

Results (Section II): Moving from regression analysis on a limited dataset to classification on a large dataset, I had new challenges to focus on. For example, computational power and time became limiting factors. To balance the computational cost with accuracy, I used five rounds of cross validation for the tuning of each hyperparameter. Additionally, I provided more limited options for

Method	Accuracy
Random Forest Classifier	94%
Support Vector Classifier (linear kernel)	85%
Decision Tree Classifier	80%
Support Vector Classifier (radial kernel)	70%
Support Vector Classifier (polynomial kernel)	50%
Neural Network	37%
Ridge Classifier	18%

Table 1: Predicting the ecoclimatic domain from the initial soil distribution data (all horizons). Each accuracy score is on the test data set and represents the optimally tuned model of the tested hyperparameters

hyperparameters than I would have liked for the neural networks, gradient boosting classifier, and support vector classifiers.

Each model listed in Table 1 is the optimal model for that method based on provided hyperparameters. Since X had size (3037, 49), it would take over an hour to perform grid search with cross validation for certain models, such as the SVCs and the neural network. Along with being the most accurate classifier (as seen in Table 1), bagging had the advantage of running extremely well on this large data set. It would be significantly faster to test many different max feature values than to tune the hyperparameters of the model. For the advantages of speed and accuracy, I would select this model for classification and would start with the random forest classifiers when expanding this work to other datasets.

Future work: If I had more time, I would also test the other datasets to see how well they characterize the ecoclimatic domains and to figure out which variables are most important for determining that. Additionally, I would use unsupervised learning on each dataset to see if the basic methods (k-means and hierarchical clustering) would recapture the ecoclimatic groupings. Both of these methods would require more computing power as well as time. In that vein, I would also try each method with more limited features and compare the accuracy to the method with the full dataset. If the accuracy was comparable, I would spend much more time testing hyperparameters for each method. Similarly, I would want to figure out the ranges and averages of different variables that

characterize each ecoclimatic region across datasets to figure out which are most distinctive in different regions. I would also want to examine the data across soil horizons to see how that affects classification accuracy. Along with classification, I would want to assess the similarities and differences between the different ecoclimatic regions. These patterns would be interesting to examine, as certain variables should be more similar based on proximity, latitude, and longitude. For example, I would expect the taiga and tundra to be more similar than either with the Pacific tropical environment, but it would be interesting to see that variation geochemically as opposed to just climatically. Additionally, I would like to explore how climate change is affecting these properties and see if there are well-documented examples of how the soil chemistry changes as a biome is exposed to highly different climate conditions.

Longer term, I would look into getting more data from more sources, particularly from other countries that may be compiling similar databases. One challenge I anticipate in this respect is ensuring consistency across chemical measurements, and I will describe this with phosphorus as an example. For my analysis, I used the Mehlich III total phosphorus results for fitting. This is because, of the methods used in the NEON dataset, it is one of the few recommended for a large range of pH values, although recent research shows it may similarly be sensitive to pH ('Phosphorus Soil Testing Methods', Penn 2018). Many phosphorus measurements only work at alkaline or acidic pH ranges, which makes them difficult to compare phosphorus across different biomes (Watson 2007). However, specific regions tend to have specific ranges or guidelines based on their method of measurement, which may not be the Mehlich III (Watson 2007). In fact, the Mehlich III is a newer test (developed in 1984) that is known for measuring higher phosphorus levels than the other methods (by as much as 6 times) (Wolf 1985, Cade-Menun 2018). Since it is newer and its results can't be directly compared to those from the Olsen or Bray extractable measurements, many labs may not use it when the Olsen and Bray measurements are baked into local regulations (Watson 2007, Wolf 1985). Globally, many different tests are used to measure phosphorus in soil, and these are difficult to compare across datasets. This is one of the advantages to the NEON data, as they conduct the same measurements across all sites.

However, when comparing different datasets, it is necessary to site average, which limits the data we have available for machine learning.

Description of Effort: Much of the work in this project went into the literature review and gaining familiarity with the data, including but not limited to understanding the different testing methods for measuring the different chemicals and the NEON specific packages required to process the data. Literature review was especially essential, as this is a field in which I lack familiarity. To establish scope and determine a reasonable research topic took a fair amount of literature searching. Once I found my topic, I then had to spend a lot of time on the NEON website learning about the data collection process, the different sites and domains, and the unique ways NEON data needs to be unzipped. After I had the data in a usable format, I was able to translate the protocol in R from the Santos paper easily into Python by following the guidelines in the paper, the github repository, and the supporting information. Replicating their results was a seamless process, but once I went to explore the relationship between soil phosphorus and pH with foliar photosynthetic traits, I hit a huge roadblock. A lot of time was spent researching 1) the different methods of phosphorus collection 2) how to handle situations with limited data. With the dataset being so small, I felt unable to confirm the null hypothesis. However, I couldn't get more and maintain consistency across all methods and sites. Since most guidance around handling limited data had far more data than I did, I decided to add a second section to my project, since there wasn't a significant result from the first part. Evaluating different models for classification required a lot of computing resources and Colab was insufficient, so I paid for a month of Colab Pro to ensure my code would run. Finding more research for the additional topic was time consuming. Overall, much more time was spent learning about the data, field, and current research, as well as waiting for code to run, than on coding.

Bibliography:

1. Aitkenhead, M.J., M.C. Coull, W. Towers, G. Hudson, and H.I.J. Black. "Predicting Soil Chemical Composition and Other Soil Parameters from Field Observations Using a Neural Network." *Computers and Electronics in Agriculture*82 (March 2012): 108–16. https://doi.org/10.1016/j.compag.2011.12.013.
2. Albrechtova, Jana, Zdeněk Seidl, Jacqueline Aitkenhead-Peterson, Zuzana Lhotáková, Barrett N. Rock, Jess E. Alexander, Zbyněk Malenovský, and William H. McDowell. "Spectral Analysis of Coniferous Foliage and Possible Links to Soil Chemistry: Are Spectral Chlorophyll Indices Related to Forest Floor Dissolved Organic C and N?" *Science of The Total Environment* 404, no. 2–3 (October 2008): 424–32. https://doi.org/10.1016/j.scitotenv.2007.11.006.
3. Cade-Menun, B.J., Elkin, K.R., Liu, C.W. et al. Characterizing the phosphorus forms extracted from soil by the Mehlich III soil test. Geochem Trans 19, 7 (2018). https://doi.org/10.1186/s12932-018-0052-9
4. Cornelissen, Johannes H., Florus Sibma, Richard S. Van Logtestijn, Rob A. Broekman, and Ken Thompson. "Leaf Ph as a Plant Trait: Species-Driven Rather than Soil-Driven Variation." *Functional Ecology* 25, no. 3 (August 12, 2010): 449–55. https://doi.org/10.1111/j.1365-2435.2010.01765.x.
5. Dijkstra, Feike A., Biao Zhu, and Weixin Cheng. "Root Effects on Soil Organic Carbon: A Double-edged Sword." *New Phytologist* 230, no. 1 (December 10, 2020): 60–65. https://doi.org/10.1111/nph.17082.
6. Kleinman, P. J. A., Sharpley, A. N., Gartley, K., Jarrell, W. M., Kuo, S., Menon, R. G., ... Skogley, E. O. (2001). Interlaboratory comparison of soil phosphorus extracted by various soil test methods. Communications in Soil Science and Plant Analysis, 32(15–16), 2325–2345. https://doi.org/10.1081/CSS-120000376
7. Maire, Vincent, Ian J. Wright, I. Colin Prentice, Niels H. Batjes, Radika Bhaskar, Peter M. van Bodegom, Will K. Cornwell, et al. "Global Effects of Soil and Climate on Leaf Photosynthetic

Traits and Rates." *Global Ecology and Biogeography* 24, no. 6 (April 7, 2015): 706–17. https://doi.org/10.1111/geb.12296.

8. Penn, Chad J., et al. "A Discussion on Mehlich-3 Phosphorus Extraction from the Perspective of Governing Chemical Reactions and Phases: Impact of Soil Ph." *MDPI*, Multidisciplinary Digital Publishing Institute, 2 July 2018, www.mdpi.com/2077-0472/8/7/106.
9. "Phosphorus Soil Testing Methods." *Cornell University - Agronomy Fact Sheet Series*, Cornell University , nmsp.cals.cornell.edu/publications/factsheets/factsheet15.pdf. Accessed 10 June 2024.
10. Santos, Fernanda, and Elizabeth Herndon. "Plant-soil Relationships Influence Observed Trends between Manganese and Carbon across Biomes." *Global Biogeochemical Cycles* 37, no. 1 (January 2023). https://doi.org/10.1029/2022gb007412.
11. Watson, Maurice, and Robert Mullen. "Understanding Soil Tests for Plant-Available Phosphorus." *Ohio State University Extension*, Ohio State University Extension, June 2007, agcrops.osu.edu/sites/agcrops/files/imce/fertility/Soil_Tests.pdf.
12. Wolf, A. M., & Baker, D. E. (1985). Comparisons of soil test phosphorus by Olsen, Bray P1, Mehlich I and Mehlich III methods. Communications in Soil Science and Plant Analysis, 16(5), 467–484. https://doi.org/10.1080/00103628509367620
13. Yi, Bo, Chaoqun Lu, Wenjuan Huang, Wenjuan Yu, Jihoon Yang, Adina Howe, Samantha R. Weintraub-Leff, and Steven J. Hall. "Resolving the Influence of Lignin on Soil Organic Matter Decomposition with Mechanistic Models and Continental-scale Data." Global Change Biology 29, no. 20 (July 13, 2023): 5968–80. https://doi.org/10.1111/gcb.16875.
14. Zhao, X., Yang, Y., Shen, H., Geng, X., and Fang, J.: Global soil–climate–biome diagram: linking surface soil properties to climate and biota, Biogeosciences, 16, 2857–2871, https://doi.org/10.5194/bg-16-2857-2019, 2019.