An Analysis into the Impact of Federal Land Protection on Environmental Quality

Research team: Lilliana Bieda, Jake Bloom, Melissa De Leon, Penelope Gordon, Richard Li, Adithya Prakash, Sohaan Shah, Saif Syed, Jahnavi Thakkar, and Samantha Zilla

Education level: Higher Secondary

Advisors: Cassie Soeffing, Peder Nelson, and Rusty Low

Affiliations: The University of Texas at Austin(NASA SEES), Austin, TX

August 5, 2025

Abstract

Protected land, like federal preserves and national parks, is often thought to be ecologically healthier than unprotected areas, but this might not actually be the truth in some cases. Protected land is often large and unmanageable, especially due to recent national park unemployment rates and cuts in funding for nature preserves nationwide. Unprotected land, on the other hand, normally has constant new environmentally friendly developments such as landscaping and green infrastructure. So this raises the question: "What are the land cover trends and environmental quality changes over time when comparing protected and unprotected areas across different locations?" The majority of our team lives in suburban areas, areas that as shown on Earth Map's Dynamic World layer, are over 50% built-up; so, with our local knowledge of our AOI's that came from collecting land cover data using GLOBE observer, we wanted to research just how different our AOI's were in terms of ecosystem health from similar protected areas. As Protection Pioneers, our research is dedicated to studying the ecosystem health, in terms of terms of plant and soil water loss, which indicates more vegetation(ET), land surface temperature(LST), photosynthesis and plant growth(GPP), and the measurement of how healthy vegetation is in an area(NDVI), of protected and unprotected areas from 2002-2022 using a coded analysis tool that will compare the metrics of the unprotected areas we live in and protected areas with similar environmental conditions like similar temperatures, elevation, and precipitation. The tool then generates downloadable line graphs comparing environmental quality metrics between the AOI and its best-matched protected site. From the research we have already done, we have found that there are distinct differences between areas of protected and unprotected land from our definition of

ecosystem health like higher plant productivity(GPP), higher averages of vegetation health(NDVI), lower land surface temperatures(LST), and more water loss(ET). In the future, we as a team hope to expand our research by collecting data at other AOI's near us. As the land cover in the United States is ever-evolving, it is important to continue monitoring the changes underway to use research like ours to create healthy environments anywhere.

Introduction/Research Question & Hypothesis

Remote sensing enables scientists to act as time travelers, and for our team, traveling into the past and future was crucial for our understanding of protected and unprotected areas. Protected areas are meant to slow down habitat loss, deforestation, and declining animal populations, but how effective is this protection? There can be a lot of arguments about how well, or how not well, protected areas can do their jobs, and this paper will compare the ecosystem health metrics of an unprotected area and a protected area related to that area in terms of plant and soil water loss(ET), land surface temperature, photosynthesis and plant growth(GPP), and the measurement of green vegetation in an area(NDVI).

Protected and unprotected areas all have their benefits and disadvantages. The purpose of this project is not to put protected areas and unprotected areas against each other; it's to understand how different they are from each other and what problems these differences could cause. Our main research question is: "What are the land cover trends and environmental quality changes over time when comparing protected and unprotected areas across different locations?" What we expect to find is that protected areas are "healthier" in terms of ecosystem health. But, with the research we do along the way to confirm this hypothesis, we're going to determine if recent conditions, like global warming and urbanization(mostly found in unprotected areas), have major beneficial or negative effects on both protected and unprotected areas. Protected areas are known for their ability to preserve ecosystems and environments, creating benefits like preventing species extinction, improving air quality, and increasing vegetation(WCMC, 2021). On the other hand, unprotected areas are seen as more negative towards local environments, causing pollution, deforestation, and urbanization(Lanzas et al., 2021). The products(either positive or negative) of both types of land come from what variables are found within their environments that affect ecosystem health, like buildings, access to sunlight, infrastructure, species populations, or even pollution.

But, indicating variables that affect ecosystem health can be difficult, as a lot of problems and endangerments of ecosystems actually come from factors that protected areas and unprotected areas can't control, like global warming and invasive species, or human-made regulations. An example of a human-made regulation that could have an effect on protected areas is that, by IUCN standards, 75% of a protected area has to be protected, and the rest can be open to the public or used for other purposes according to the IUCN(*Template:IUCN Categories - OpenStreetMap Wiki, 2022*). This regulation is interesting because protected areas were created to keep certain areas of nature away from people, for their protection. And specifically, this point is related to the United States, as national parks have become tourist attractions that attract as many people as you would see in a city. Funding is one of the main issues with modern protected areas because these large areas require lots of management, but most preserves and parks don't have the money to fund it. By recording the ecosystem health of protected and unprotected areas, future problems like global warming can be prevented and "solved".

Everyone on the Protection Pioneers team mostly resides in suburban areas that are nearby by federal preserves and parks, so there is a good mix of both worlds where we live. Our locations was one of the main inspirations as to why we decided to record the differences between protected areas and suburban/urban areas, as our areas have become more built-up in recent years with more people coming to live in them; from these urbanization trends, we have come to realize that our suburban areas are starting to become less environmentally friendly due to factors like infrastructure and urbanization, but at the same time, our areas are implementing preventions like green infrastructure and conservation programs. All of these factors, like urbanization or green infrastructure, are ecosystem health "variables" that help determine why an environment is improving or declining, no matter if it's protected or unprotected.

The WDPA (World Database on Protected Areas) is one of the best places to get an understanding of protected areas because the WDPA website showcases a database of all the protected areas in the world and provides yearly information about them(even if a protected area near you isn't that well known, the WDPA will have it listed on their site)! The WDPA website offers a great introduction to protected areas and how their vegetation and wildlife have changed over the years, which allows for research projects like ours to create beneficial and relevant studies about protected areas. Furthermore, studies from researchers like Claudia L. Gray, who reported in the scientific journal Nature, that species and vegetation are much higher in protected areas, are essential to understanding what is effective in both protected and unprotected areas, and what is not. From knowing how environments are affected by variables within their ecosystems, researchers can create future projects and programs that can benefit environments in any area. Furthermore, while 15% of the Earth is represented by protected areas, the same article from Nature references the point that only 22% of them have effective management (management is an example of a key ecosystem variable) to allow for linear growth; and, since there has been an increase in tourism over the years, this linear growth is starting to slow down, and in some cases, decline(Gray et al., 2016).

Methodology

In order to complete this analysis, every team member started by selecting an Area of Interest (AOI) and received an assigned 3km by 3km grid surrounding that area to collect data. These grids contain 37 specific locations, and each member travelled to each location and recorded a GLOBE Observer Land Cover Observation, for a total of 37 per person and 370 between the entire group. Each observation consisted of a land cover description, where ground moisture, tree cover percentage, water content, along with 6 photographs, facing up, down, north, east, south, and west. Whilst collecting GLOBE data, members also emphasized observing the forms of land use in their AOI in order to draw connections and correlations between the usage of land and the environmental health factors.

To draw the most accurate comparisons between protected and unprotected areas, we developed an automated program that identifies the most ecologically similar protected area to the given unprotected AOI. The tool then generates downloadable line graphs comparing environmental quality metrics between the AOI and its best-matched protected site over the years 2002-2022. Once each user completes their comparison, we combine all results for an aggregate analysis. From this, we test statistical significance for differences in environmental conditions between protected and unprotected areas, temporal trends (i.e., whether conditions improve over time with development), and differences in variability (volatility) between protected and unprotected areas.

We began by downloading a CSV dataset of federally protected areas in the United States from the WDPA. To reduce our tool's computational strain, we cleansed the protected land dataset to only include what we needed using a Python script in Colab: removing marine sites (since no members' Areas of Interest were aquatic), eliminating duplicates, and excluding areas under 80 km^2 to avoid very small or fragmented sites — though this threshold exceeded the actual resolution needs of our analysis. This preprocessing reduced the dataset from roughly 51,000 entries to 1,355. We also trimmed the original 31 columns down to six: name, year of designation, managing organization, latitude, longitude, and area (km^2).

We then appended our protected land dataset to include columns for average annual temperature (°C, 1991–2021), total annual precipitation (mm), and elevation (m)—biome-defining metrics used to assess which protected land site is most ecologically similar to the user's AOI, and thus most suited for comparison over time. These values were retrieved using a Python script in Colab that pulled data from Open Meteo API, a free historical climate data service. While this step could have been done in real time within the tool, making 3 API requests for each of the 1,350+ sites would have been a long process, so locally storing it ahead of time was simply better for efficiency & user convenience.

With a comprehensive database of protected lands in the US and their respective relevant climate data, we developed a Python-based web application, coded through GitHub Codespaces, and deployed via Streamlit, which you could find here: sees-protection-tool.streamlit.app. Upon arriving at the landing screen, users are prompted to upload a CSV file representing their AOI so the program can extract their centroid coordinates. The model then retrieves climate and elevation data for those coordinates, using the same coding logic & data source (Open Meteo) as were used for the protected areas.

Once we gather the climate and elevation data from the user's location, we compare it to all the protected areas in our dataset to find the most similar one. To ensure a fair comparison, we first convert each variable – including temperature, rainfall, and elevation – into z-scores, which show how typical or unusual each value is compared to the rest of the dataset. This standardization prevents any single factor from dominating the results just because it uses larger numbers (such as rainfall or elevation).

To actually find the best match, we use the Euclidean Nearest Neighbors algorithm provided by scikit-learn – "Euclidean" simply meaning it looks for the straight-line distance between two points, and "Nearest Neighbors" meaning it ranks which protected areas are closest in terms of overall ecological similarity. While more complex models like decision trees, clustering, or regression could be used, they assume patterns or groupings in the data. Since our use case isn't to predict or categorize, we picked Euclidean Nearest Neighbors. The result is a ranked list of protected areas, with the top match selected for deeper comparison.

Specifically, the app then generates line graphs comparing the user's location to the best similar protected area from 2002–2022 across the following environmental indicators: NDVI (vegetation health), ET (water use), LST (surface temp), and GPP (plant growth). These were sourced from Google Earth Engine's MODIS (006) dataset at a 3km x 3km resolution to match the AOI size. A downloadable CSV of the data used on the graphs is also provided, intended for group-wide aggregated analysis.

To assist in aggregate analysis for numerous metrics, we used a Python script in Google Colab to compile several CSV files and generate graphs, as discussed throughout the following paragraph. Group Averages, which are average yearly values for each group (protected vs. unprotected) with 95% confidence intervals, provided a clear visualization of the general trends between protection statuses. Year-by-Year Boxplots compare distribution values for protected and unprotected areas each year, allowing for the visualization of the data spread. Standard deviation values for each group were also calculated on a yearly basis in order to track variability and margin of error over time, as increases in variability could signify potential environmental instability and decreases in variability could signify environmental consistency. We employed a mixed-effects model to account for inherent differences between sites, allowing us to assess whether the observed differences were consistent for each year and site. Finally, we used tests such as Paired t-tests, Wilcoxon tests, and Levene's test to evaluate the statistical significance of differences and variations between the different levels of protection.

Discussion/Future Recommendations

Over the course of our research, we have taken several steps to reduce bias and improve the reliability of our results. However, if replicated, our research could be improved by incorporating more metrics into our tool when comparing our AOIs and an ecologically similar protected area. We were able to draw conclusions based on the trends we observed from the Normalized Difference Vegetation Index (NDVI), Evapotranspiration (ET), Land Surface Temperature (LST), and Gross Primary Productivity (GPP) graphs over time, but more metrics would create new opportunities to observe different aspects of an area. Additional metrics, for example, soil moisture information, could provide insight into countless other topics and more specific research questions regarding the growth of vegetation and more. That would increase the quantity of outputs, but there is always room to improve the quality of outputs. While using the tool described in our methodology above, certain statistical tests were used and we were not able to meet all the criteria needed for these tests. This should be noted when using our results and conclusions. If our investigation were to be repeated, the use of a larger and more random sample size would help meet these criteria and improve the reliability of our results.

Although there are ways to recreate and improve our research, there are also many interesting ways to expand upon it and explore other topics. Our research ended up being primarily focused on the vegetation of an area and its health, along with surface temperature. There are a plethora of other ways to view a protected or unprotected area. The biodiversity of an area could be explored in terms of just plants or animals, too. Soil and weather can also be investigated and used to give people a more complete understanding of how protected status affects an ecosystem as a whole.

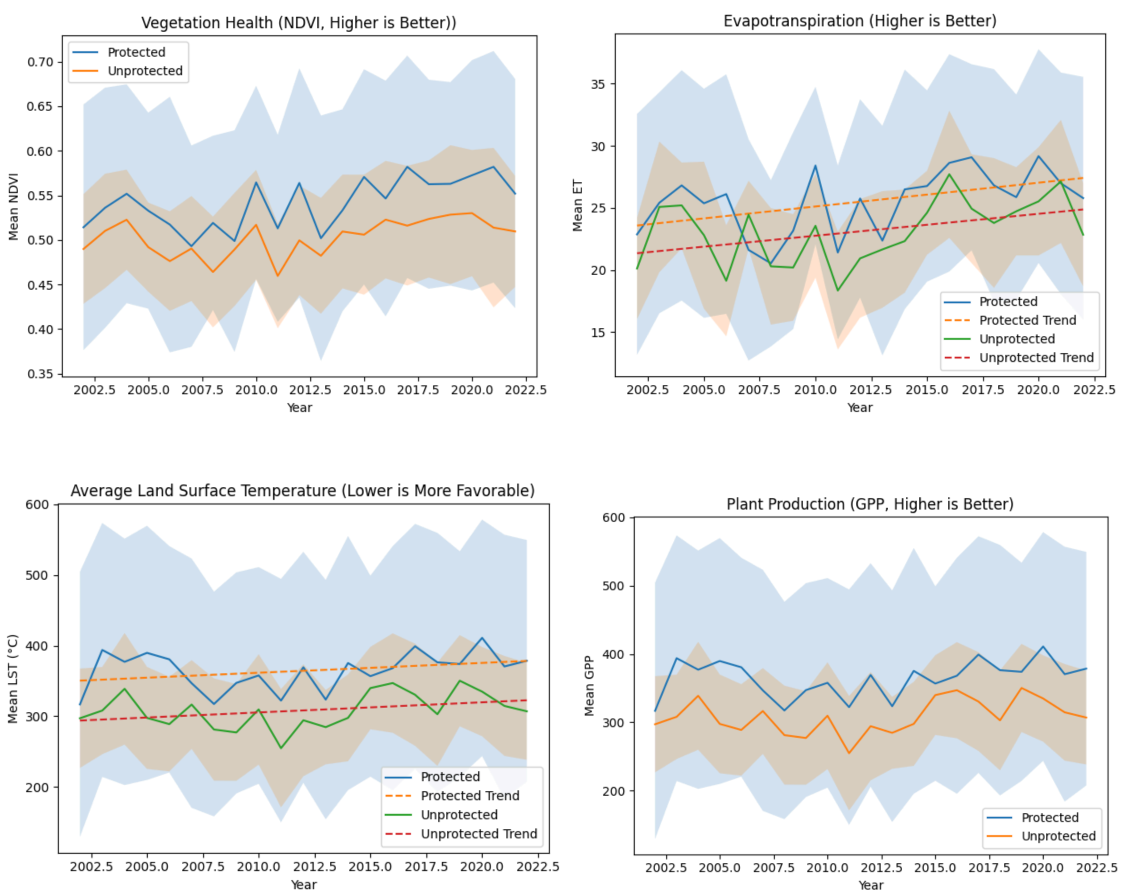
In addition, the effects of proximity to anthropogenic development have been and should be further investigated. It is easy to generalize protected areas and think of them as homogeneous blocks of land, but there are nuances within these places, specifically the edges and central areas. Multiple trends occur around the edges of a protected area, and those trends could be further explored as an extension to our research. Research has been done on this topic already as seen in the *Effectiveness of protected areas edges on vegetation greenness, cover and productivity on the Tibetan Plateau, China* article. They describe how edges of protected areas still show the benefits to vegetation, but trends are not as strong as those in central areas (Hua et al., 2022). The spillover effect, when positive effects from protected areas carry over into non-protected areas nearby, is also noted (Hua et al., 2022).

The impacts of our research and other research like it could have valuable effects on protected areas, more specifically, National Parks. Support for or criticisms of protected areas could affect public awareness and backing of institutions like the National Parks Service. As always, funding could also be affected by research and public opinion. Aside from that, information about the edges of protected areas can help with the design of protected areas to maximize their effects. The degree to which fragmentation of protected areas should be avoided can be determined by the severity of the effects of the edges of protected areas. Spillover effects can also be incorporated into the design of protected areas to help surrounding communities.

Results

Environmental Quality Differences

Protected areas consistently outperformed unprotected areas across all ecological metrics. Mean and median values were higher for NDVI (vegetation health), GPP (photosynthetic productivity), and ET (evapotranspiration), while Land Surface Temperature (LST)—a proxy for development—was significantly lower. These differences were confirmed as highly statistically significant, using both paired t-tests and Wilcoxon signed-rank tests ($\alpha = 0.05$), indicating that, at face value, protected lands consistently exhibit better environmental conditions.



Graphs of mean ecological indicators for protected vs. unprotected sites from 2002-2022, with 95% confidence intervals.

Note: The protected & unprotected lines largely mirror each other in shape, suggesting each set of locations experienced similar external environmental pressures (e.g., extreme weather)

Trends in Change over Time

Despite superior average values in protected areas, both site types improved at largely similar rates from 2002–2022, suggesting protection may have not been the primary driver of higher averages. To isolate the effect of protection status, we used mixed-effects models that controlled for site-level variation. Results showed no statistically significant influence of protection on NDVI (p = 0.421), GPP (p = 0.847), or ET (p = 0.901)—not only in terms of changes over time, but also in overall magnitude. Only Land Surface Temperature (LST) remained significantly lower in protected areas (p = 0.033). This implies that protected sites may have started with better baseline conditions for NDVI, GPP, and ET, which protection merely preserved, explaining the higher averages. In contrast, the significant LST reduction likely reflects protection's active role in limiting development, which typically raises surface temperatures.

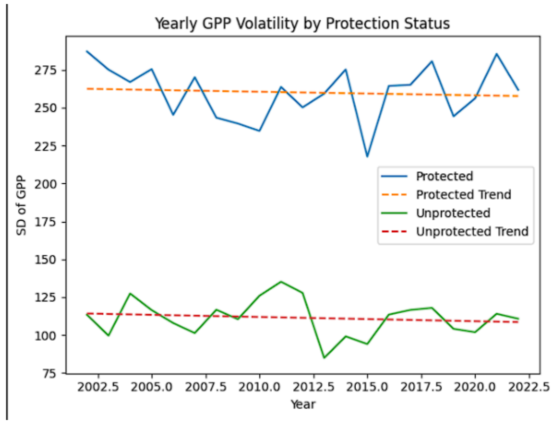
```
Paired NDVI t-test: t=3.741, p=0.000
Paired NDVI Wilcoxon: stat=7983.500, p=0.000
                   Mixed Linear Model Regression Results
==========================================================================
Model:                     MixedLM        Dependent Variable:        NDVI
No. Observations:          420            Method:                    REML
No. Groups:                10             Scale:                     0.0079
Min. group size:           42             Log-Likelihood:            382.5510
Max. group size:           42             Converged:                 Yes
Mean group size:           42.0
--------------------------------------------------------------------------
                                  Coef.  Std.Err.   z    P>|z| [0.025 0.975]
--------------------------------------------------------------------------
Intercept                         -5.039    2.036 -2.475 0.013 -9.030 -1.049
C(protection)[T.Unprotected]       2.278    2.879  0.791 0.429 -3.364  7.920
year                               0.003    0.001  2.740 0.006  0.001  0.005
C(protection)[T.Unprotected]:year -0.001    0.001 -0.805 0.421 -0.004  0.002
Group Var                          0.008  170.765
Group x year Cov                  -0.000    0.083
year Var                           0.000    0.000
==========================================================================


Levene test for equal NDVI variances: stat=87.503, p=0.000
```

Mixed-effects model showing that protection status did not significantly influence NDVI trends over time (interaction p = 0.421), suggesting limited impact of protection alone on ecological health.

Variability Differences

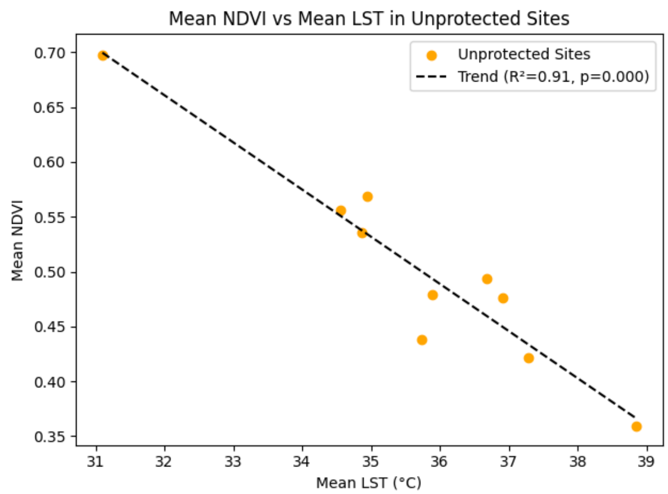
Protected areas consistently exhibited higher variability across all environmental metrics compared to unprotected areas. This pattern held across box plots, standard deviation graphs, and Levene's test, indicating significantly higher environmental fluctuation in protected lands. Mixed-effects models isolating the impact of protection status on variability further confirmed this: even after removing outliers, protection was a significant predictor of strongly increased standard deviation in GPP (p = 0.007), NDVI (p = 0.007), and LST (p = 0.018). ET showed a similar trend (p = 0.120) but was not significant. These results suggest that unprotected areas—likely influenced by urbanization, land regulation, and infrastructure designed to resist environmental stressors—may buffer environmental changes more effectively. In contrast, protected areas, which lack such artificial defenses, remain exposed to unmediated forces like extreme weather events, leading to greater ecological volatility.



Example of consistent, higher volatility.

Role of Development and Metrics

To isolate the role of land development influencing ecological averages across unprotected sites, we plotted mean LST—a proxy for heat-inducing development—against each ecological metric. The results show a strong negative correlation between LST and NDVI ($R^2 = 0.91$, $p < 0.05$), suggesting that development (or at least that which induces heat) is significantly associated with degrading vegetation health, which may explain the lower NDVI averages in unprotected sites. However, when comparing LST to GPP (plant production) and ET (water use), a significant relationship was not found.



Limitations & Notes: sample size of ten matched pairs (one missing ET & GPP values), and use of a single 3 km×3 km sample per protected area, which may not capture full site.

Conclusions & Closing Thoughts

Our analysis revealed consistent ecological advantages in protected areas: higher average vegetation health (NDVI), plant productivity (GPP), evapotranspiration (ET), and lower land surface temperatures (LST). These statistically significant differences initially suggested that protected areas were inherently superior. However, further analysis complicates that view. After isolating protection status using mixed-effects modeling, only LST remained significantly affected, both in magnitude and change over time. This implies the elevated NDVI, GPP, and ET may stem from preexisting favorable conditions (i.e. selection bias), not protection itself. In contrast, LST's influence by protection is likely due to protection's active role in limiting development, which tend to increase surface temperatures. These explanations, while plausible, remain speculative and require further testing.

Equally compelling was how protected areas had more – not less – environmental volatility, even after removing outliers that may have skewed results. The fact this happened despite our efforts to control for biome & climate, suggests unprotected areas, perhaps through land use controls, were able to regulate their environment in a way that would make them less susceptible to the (generally) same external pressures (an assumption made based off of how the line graphs tended to mirror each other in shape).

These findings suggest a fascinating conclusion: Protection status does not universally enhance all ecological metrics; instead, its most tangible effect appears to be in buffering against urbanization-driven heat increases. Meanwhile, unprotected areas, perhaps due to active land management, demonstrate a surprising degree of environmental stability. While limited by a small sample size (10, with one site missing GPP and ET values) and insufficient spatial resolution to capture full protected area boundaries, this complicates the narrative that "protection equals better" and underscores the need for more targeted analyses of how specific conservation strategies affect ecological outcomes.

References

ArcGIS online. arcgis.com. (n.d.). https://www.arcgis.com/index.html

*Interns accessed ArcGIS to learn more information about their Areas of Interest and GLOBE Observations.

Hua, T., Zhao, W., Cherubini, F., Hu, X., & Pereira, P. (2022). Effectiveness of protected areas edges on vegetation greenness, cover and productivity on the Tibetan Plateau, China*.* *Landscape and Urban Planning*, 224, 104421. https://doi.org/10.1016/j.landurbplan.2022.104421

WCMC. (2021, February). *Successful protected areas are a matter of quality, not just quantity.* UNEP-WCMC.https://www.unep-wcmc.org/en/news/successful-protected-areas-are-a-matter-of-quality--not-just-quantity

Lanzas, M., Hermoso, V., Morán-Ordoñez, A., Regos, A., Bota, G., & Brotons, L. (2021). The value of unprotected land for future conservation efforts under dynamic conditions. *Biological Conservation*, 261, 109232. https://doi.org/10.1016/j.biocon.2021.109232

Template: IUCN Categories - OpenStreetMap Wiki. (2022). Openstreetmap.org. https://wiki.openstreetmap.org/wiki/Template:IUCN_Categories

Gray, C. L., Hill, S. L. L., Newbold, T., Hudson, L. N., Börger, L., Contu, S., Hoskins, A. J., Ferrier, S., Purvis, A., & Scharlemann, J. P. W. (2016). Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nature Communications*, 7(1). https://doi.org/10.1038/ncomms12306

Earth Map Land Cover Analysis Tool. Earth Map. (n.d.). https://earthmap.org/?map=%7B%22center%22%3A%7B%22lat%22%3A0%2C%22lng%22%3A0%7D%2C%22zoom%22%3A3%2C%22mapType%22%3A%22roadmap%22%7D

*Earth Map is a tool provided to researchers to analyze land cover data layers. Many layers were used with this tool.

GLOBE Observe Land Cover Data Observations. GLOBE.gov. (n.d.). https://observer.globe.gov/

*Interns used the GLOBE Observer App to make land cover photo observations. Each student used this app throughout June to observe their Areas of Interest (AOIs).

Syed, S., Shah, S., Prakash, A. (2025, July). *Protection Pioneers Land Cover Tool*. SEES. https://sees-protection-tool.streamlit.app/

**Other interns helped create and publish the tool; Saif is credited with the main creation of the tool. Sohaan Shah and Adithya Prakash also assisted in the creation of this tool.*

Acknowledgements

We would like to credit each *Protection Pioneers* researcher who worked diligently throughout the research project, Lilliana Bieda, Jake Bloom, Melissa De Leon, Penelope Gordon, Richard Li, Adithya Prakash, Sohaan Shah, Saif Syed, Jahnavi Thakkar, and Samantha Zilla. We would also like to thank Peer Mentor and past SEES intern, Noah Peralez, for providing guidance as we established our research topic. Our work would not have been possible without the guidance of the Earth Systems Explorers lead mentors, Andrew Clark, Brianna Lind, Rusty Low, Peder Nelson, Erika Podest, and Cassie Seoffing. Their leadership and knowledge ensured we were successful in our citizen science endeavours, and we're forever grateful to have worked with such inspiring figures. Lastly, thank you to NASA, SEES, and CSR for making our research possible with their support, resources, and funding.

Badge Descriptions

I Work With Satellite Data

Using tools like Earthmap.org, along with other features in the World Database on Protected Areas filter, we incorporated NASA satellite data throughout our entire research to analyze land behavior in both federally protected and unprotected areas. Using these satellite-derived layers granted us the ability to analyze factors such as temperature, vegetation health, and plant growth :all key factors in determining ecological conditions across our AOIs.

I Am A Data Scientist

We conducted in-depth analysis using both data from GLOBE database and satellites according to our AOIs. We developed and coded a custom tool that explores the intricacies of our research question. By being able to analyze key environmental indications of ecological health according to soil and plant water loss, photosynthesis rates, and vegetation health, we were able to view the ecological effects based on land protection status. We also discussed data limitations such as processing time and the potential exclusion of useful sites, but results still made it possible for us to make inferences about environmental health and possible future trends.

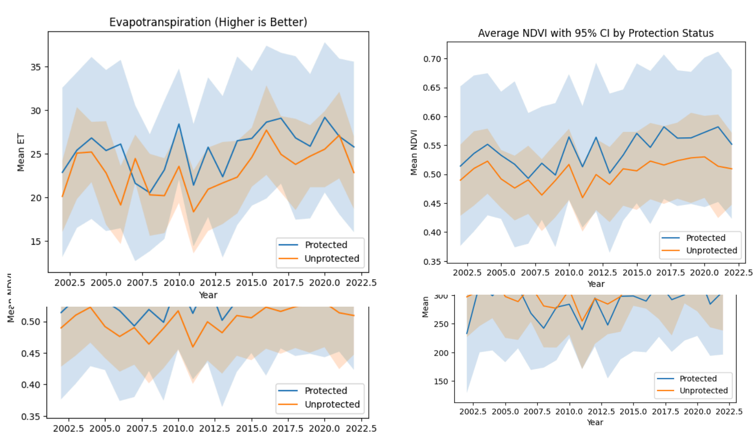
Earth System Scientist

By analyzing how the atmosphere, biosphere, and pedosphere(soil) interact in both federally protected and non-protected land means our project explores the interconnectedness of earth systems. We applied several GLOBE protocols and visualized how one system change (like reduced vegetation) can affect others (like soil moisture or temperature). And by examining this data collectively, we can highlight the importance of maintaining ecosystem balance in federally protected areas.

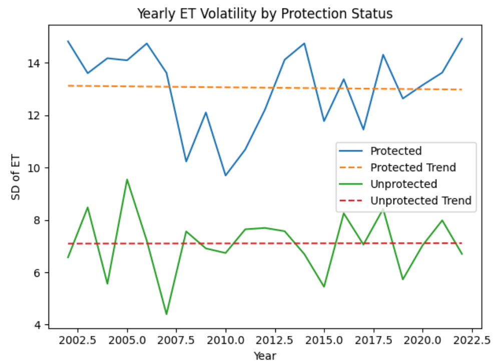
Appendix

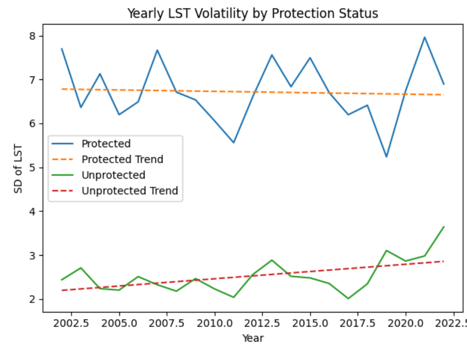
Aggregate Graphs

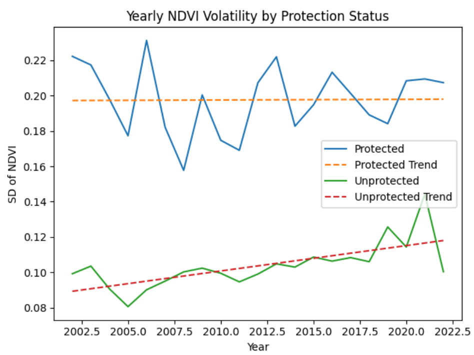
Stability

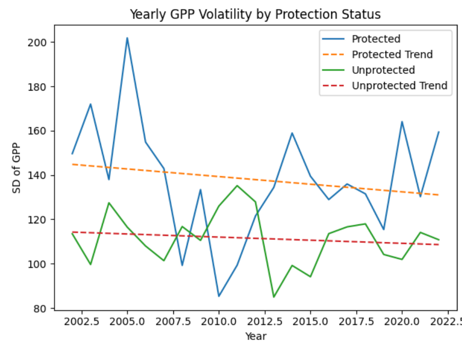


Graphs

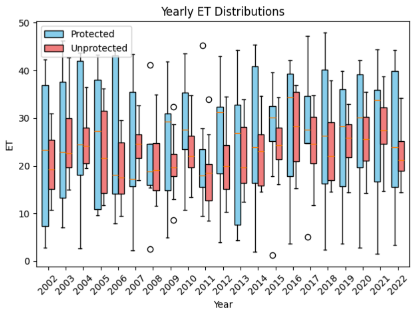


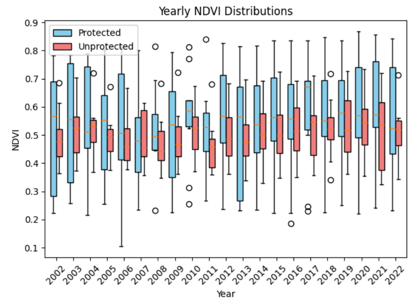


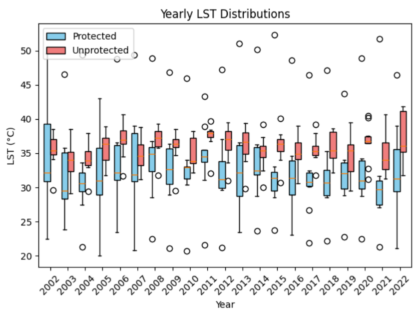


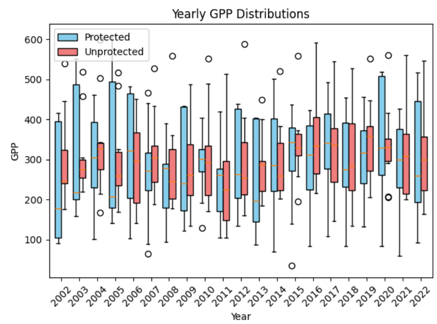


Distribution Graphs

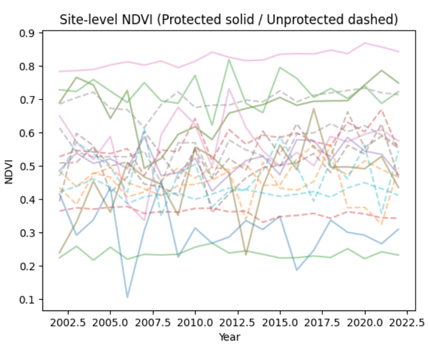


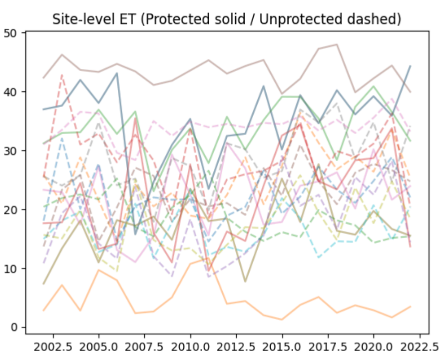


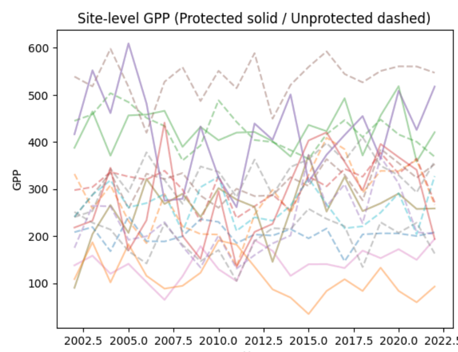




Comparison Graphs







Code

```
                 Mixed Linear Model Regression Results
=====================================================================
Model:                    MixedLM       Dependent Variable:     SD_GPP
No. Observations:         18            Method:                 REML
No. Groups:               10            Scale:                  194.5956
Min. group size:          1             Log-Likelihood:         -71.7508
Max. group size:          2             Converged:              Yes
Mean group size:          1.8
---------------------------------------------------------------------
                              Coef.  Std.Err.   z    P>|z|  [0.025 0.975]
---------------------------------------------------------------------
Intercept                     63.547    6.931  9.168 0.000  49.962 77.132
C(protection)[T.Unprotected] -18.605    6.855 -2.714 0.007 -32.040 -5.170
Group Var                    205.153   15.796
=====================================================================
```

```
               Mixed Linear Model Regression Results
===================================================================
Model:                   MixedLM       Dependent Variable:      SD_NDVI
No. Observations:        20            Method:                  REML
No. Groups:              10            Scale:                   0.0003
Min. group size:         2             Log-Likelihood:          38.9402
Max. group size:         2             Converged:               Yes
Mean group size:         2.0
-------------------------------------------------------------------
                             Coef.  Std.Err.   z    P>|z| [0.025 0.975]
-------------------------------------------------------------------
Intercept                     0.066    0.009  7.622 0.000  0.049  0.083
C(protection)[T.Unprotected] -0.021    0.008 -2.687 0.007 -0.036 -0.006
Group Var                     0.000    0.023
===================================================================
```

```
               Mixed Linear Model Regression Results
===================================================================
Model:                    MixedLM      Dependent Variable:     SD_ET
No. Observations:         19           Method:                 REML
No. Groups:               10           Scale:                  0.7621
Min. group size:          1            Log-Likelihood:         -31.4728
Max. group size:          2            Converged:              Yes
Mean group size:          1.9
-------------------------------------------------------------------
                             Coef.  Std.Err.   z    P>|z| [0.025 0.975]
-------------------------------------------------------------------
Intercept                     5.254    0.512 10.262 0.000  4.250  6.257
C(protection)[T.Unprotected] -0.635    0.408 -1.556 0.120 -1.436  0.165
Group Var                     1.716    1.625
===================================================================
```

```
Paired NDVI t-test: t=3.741, p=0.000
Paired NDVI Wilcoxon: stat=7983.500, p=0.000
                 Mixed Linear Model Regression Results
=====================================================================
Model:                      MixedLM        Dependent Variable:     NDVI
No. Observations:           420            Method:                 REML
No. Groups:                 10             Scale:                  0.0079
Min. group size:            42             Log-Likelihood:         382.5510
Max. group size:            42             Converged:              Yes
Mean group size:            42.0
---------------------------------------------------------------------
                                   Coef.  Std.Err.   z    P>|z| [0.025 0.975]
---------------------------------------------------------------------
Intercept                         -5.039    2.036 -2.475 0.013 -9.030 -1.049
C(protection)[T.Unprotected]       2.278    2.879  0.791 0.429 -3.364  7.920
year                               0.003    0.001  2.740 0.006  0.001  0.005
C(protection)[T.Unprotected]:year -0.001    0.001 -0.805 0.421 -0.004  0.002
Group Var                          0.008  170.765
Group x year Cov                  -0.000    0.083
year Var                           0.000    0.000
=====================================================================


Levene test for equal NDVI variances: stat=87.503, p=0.000
```

```
Paired ET t-test: t=2.911, p=0.004
Paired ET Wilcoxon: stat=7087.500, p=0.012
                     Mixed Linear Model Regression Results
=================================================================================
Model:                      MixedLM         Dependent Variable:        ET
No. Observations:           399             Method:                    REML
No. Groups:                 10              Scale:                     46.0942
Min. group size:            21              Log-Likelihood:            -1348.9739
Max. group size:            42              Converged:                 Yes
Mean group size:            39.9
---------------------------------------------------------------------------------
                                    Coef.   Std.Err.   z    P>|z|  [0.025   0.975]
---------------------------------------------------------------------------------
Intercept                         -359.402  164.120 -2.190 0.029 -681.071 -37.734
C(protection)[T.Unprotected]        28.061  226.195  0.124 0.901 -415.272 471.394
year                                 0.191    0.082  2.344 0.019    0.031   0.351
C(protection)[T.Unprotected]:year   -0.015    0.112 -0.134 0.893   -0.235   0.205
Group Var                           47.543 6096.846
Group x year Cov                    -0.050    3.026
year Var                             0.000    0.002
=================================================================================


Levene test for equal ET variances: stat=85.763, p=0.000
```

```
Paired NDVI t-test: t=2.789, p=0.006
Paired NDVI Wilcoxon: stat=7028.000, p=0.010
                     Mixed Linear Model Regression Results
==================================================================================
Model:                      MixedLM         Dependent Variable:        GPP
No. Observations:           399             Method:                    REML
No. Groups:                 10              Scale:                     13740.5310
Min. group size:            21              Log-Likelihood:            -2475.4074
Max. group size:            42              Converged:                 Yes
Mean group size:            39.9
----------------------------------------------------------------------------------
                                    Coef.    Std.Err.    z    P>|z|   [0.025    0.975]
----------------------------------------------------------------------------------
Intercept                         -2439.568   2850.550 -0.856 0.392 -8026.543 3147.408
C(protection)[T.Unprotected]       -139.729   3917.849 -0.036 0.972 -7818.571 7539.113
year                                  1.388      1.417  0.980 0.327    -1.389    4.165
C(protection)[T.Unprotected]:year     0.047      1.947  0.024 0.981    -3.769    3.864
Group Var                           115.973 358420.526
Group x year Cov                      0.664    178.246
year Var                              0.005      0.089
==================================================================================


Levene test for equal NDVI variances: stat=42.345, p=0.000
```

```
               Mixed Linear Model Regression Results
===================================================================
Model:                   MixedLM       Dependent Variable:      SD_LST
No. Observations:        18            Method:                  REML
No. Groups:              10            Scale:                   0.1866
Min. group size:         1             Log-Likelihood:          -13.4839
Max. group size:         2             Converged:               Yes
Mean group size:         1.8
-------------------------------------------------------------------
                             Coef.  Std.Err.   z    P>|z| [0.025 0.975]
-------------------------------------------------------------------
Intercept                     2.167    0.185 11.728 0.000  1.805  2.529
C(protection)[T.Unprotected] -0.512    0.216 -2.365 0.018 -0.935 -0.088
Group Var                     0.060    0.305
===================================================================
```

```
                     Mixed Linear Model Regression Results
=================================================================================
Model:                      MixedLM         Dependent Variable:        LST
No. Observations:           420             Method:                    REML
No. Groups:                 10              Scale:                     9.5252
Min. group size:            42              Log-Likelihood:            -1092.7489
Max. group size:            42              Converged:                 Yes
Mean group size:            42.0
---------------------------------------------------------------------------------
                                    Coef.   Std.Err.   z    P>|z|  [0.025   0.975]
---------------------------------------------------------------------------------
Intercept                          165.575   70.772  2.340 0.019   26.864 304.286
C(protection)[T.Unprotected]      -213.273  100.077 -2.131 0.033 -409.421 -17.124
year                                -0.066    0.035 -1.877 0.061   -0.135   0.003
C(protection)[T.Unprotected]:year    0.107    0.050  2.161 0.031    0.010   0.205
Group Var                            9.734 9834.866
Group x year Cov                    -0.011    4.851
year Var                             0.000    0.002
=================================================================================


Levene test for equal LST variances: stat=44.671, p=0.000
```

Raw Aggregate Data:

https://drive.google.com/file/d/1TXb66fivQ9NqJtRq7aRmkxl3cL9nljvN/view?usp=sharing

Protected Lands Dataset:

https://github.com/SaifSyed08/protected-lands-app/blob/main/protected_lands.csv