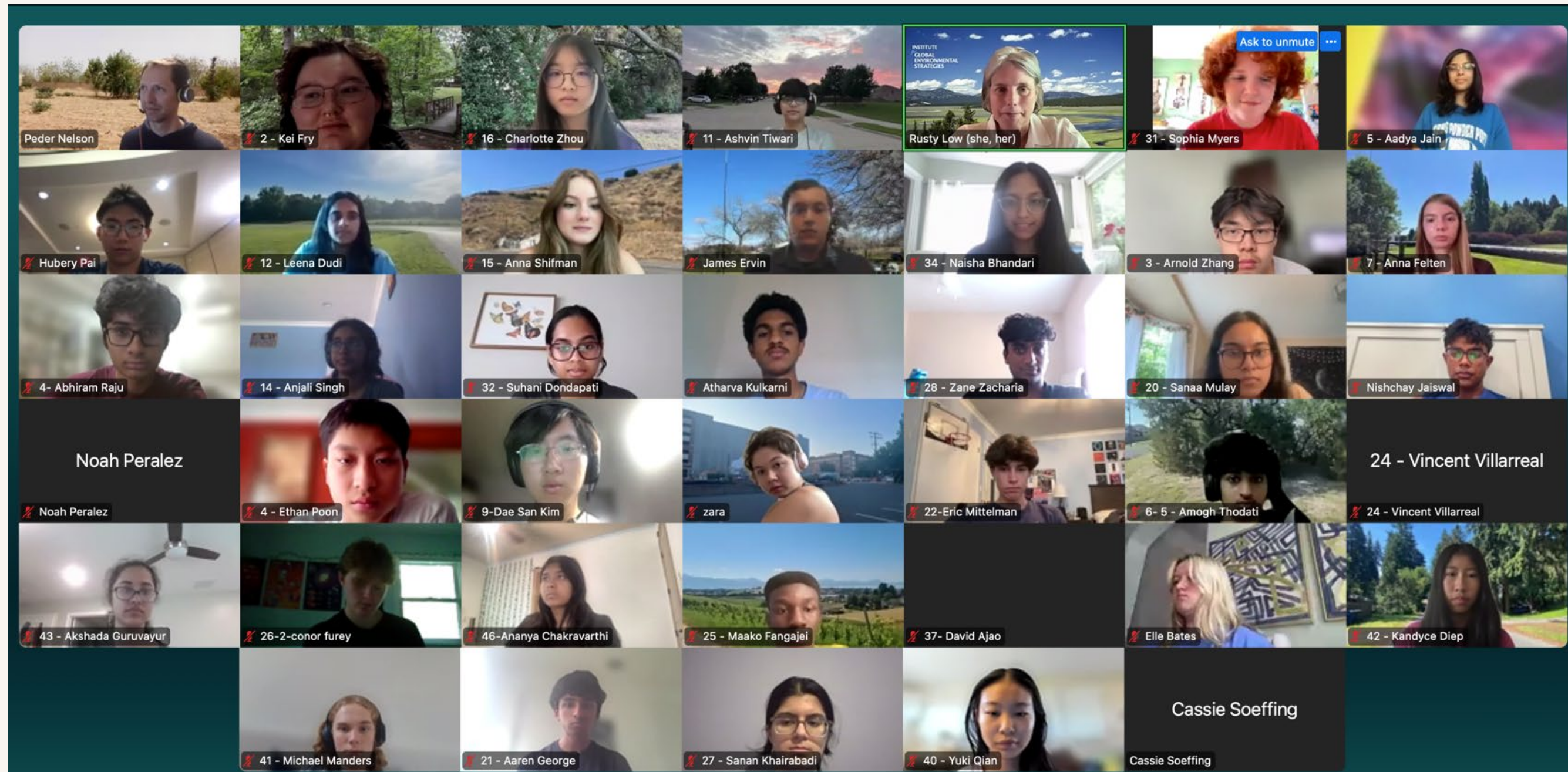


Cassie Soeffing
Cassie_Soeffing@strategies.org



Earth System Explorers (ESE) 2024



GLOBE takes students on a real-world science adventure, transforming them from observers to active researchers. Our team will share how the SEES Earth System Explorers engaged students virtually across the entire data life cycle. We'll explore how **students collected environmental data, analyzed it alongside scientists, visualized their findings, and contributed to GLOBE global datasets.**

GLOBE and NASA data support critical thinking, open science, and open data.

Earth System Explorers framework

Individual Identity

Interns enter with varying interests, skills and knowledge

Team Identity

Team Science and Leadership: foster an ethos of science through intentional design

Science Identity

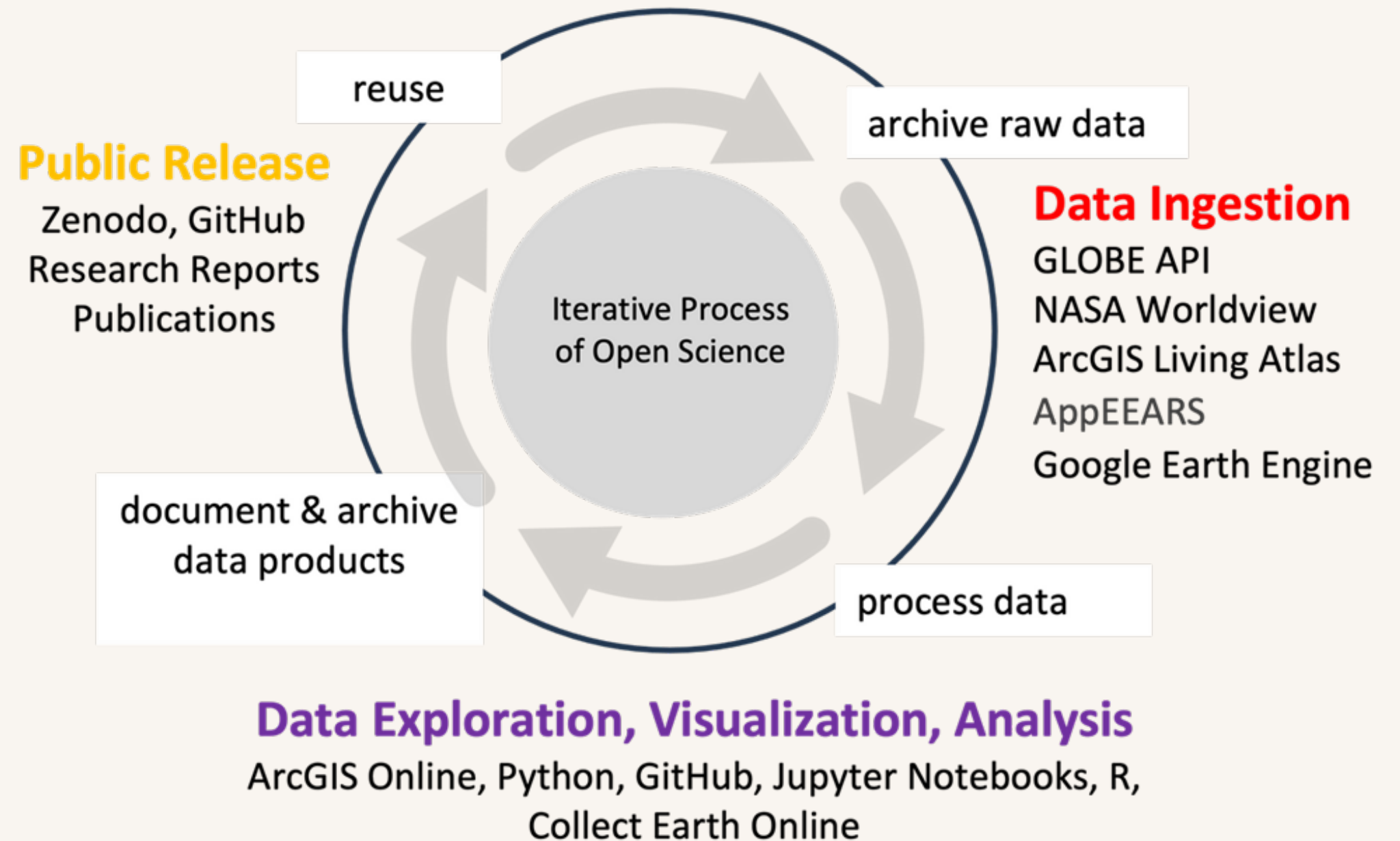
Open Science: learn about and practice open science (FAIR data)

Community Identity

Exit internship with expanded skills, knowledge, and community involvement

Field Data Collection

Using GLOBE Observer Mobile app



Growth of the Earth System Explorers

Mosquito Mappers

2018

- Cross-collaboration by two SciAct Teams: SEES (UT Texas Austin), and NESEC
- A citizen science data collection experience for highly-qualified SEES candidates.

73 interns

2019

- A month-long webinar series
- Continued mosquito-focused research component.

65 interns

2020

- Optional pre-internship modules offered by SEES (Python Coding & Remote Sensing)
- Added Land Cover and changed from 4-8 weeks.
- **Added GLOBE Observer Land Cover scientist to the core team to incorporate Adopt a Pixel and a strong Land Cover research focus in fieldwork and intern research.**

Our new nested sampling framework enables the practical application of quantitative and statistical analysis of GLOBE Observer-obtained data. This framework links remote sensing land cover imagery with field-based observations, providing a comprehensive and practical approach to data analysis.

74 interns

40 IVSS submissions

13 earned 4 Badges

Earth System Explorers

2021

Reflect interdisciplinary focus: mosquitoes, land cover, NASA Earth observations.

107 interns

28 IVSS submissions

14 earned 4 Badges

2022

- Pilot Team Science module
- All interns are required to participate in team projects.
- Used Mighty Networks community engagement platform
- Interns were encouraged to submit blogs posted on GLOBE Mission Mosquito.
- Created badges for completing internship components

98 interns

23 IVSS submissions

21 earned 4 Badges

2023

- Added Open Science concentration
- Added Data Scientist to core mentor team
- Peer Mentor training developed.
- Internship activities are connected to badges earned over the course of the summer: Adopt-a-Pixel, Mosquito Habitat Mapper, Science of Team Science, and Earth System Explorer

47 interns

11 IVSS submissions

7 earned 4 Badges

2024

Past research teams have presented at GLOBE Observer Connect and Trees Around the GLOBE Student Research Campaign.



Various Research Project Titles

2023

Comparing **AI** Algorithms and Their Classification Abilities Utilizing **CNN** to Identify the Environment Where Each Mosquito Genera Thrives

Enhancing Citizen Science Building a **Web-Based** Platform for Optimizing the Area of Interest Data Collection Project
Case Study of Hurricane Harvey (2017)

Analyzing the Correlation Between Terrain and the Characteristics of Wildfires

Uniting Deep Learning and Citizen Science for Automatic Land Cover Classification

Analyzing the Effects of Urbanization on Mosquito Population Dynamics Using Citizen Science Data

Exploring the Relationship Between Land Cover Classifications and Urban Heat Island Intensity

A **Random Forest Analysis** of Remote Sensing Driven Mosquito Habitat Prediction In West Africa

Influence of Tree Density Upon the Spatiotemporal Distribution of Mosquito Populations

Urbanization in Rural Towns - A Case Study of West Columbia, TX

2022

Region-Based **Modeling** of Atmospheric and GLOBE Data to predict Mosquito Hotspots

Analyzing the Effect of Hurricanes on Mosquito Population Patterns.

Forecasting West Nile Virus Infections: A **Machine-Learning** Approach to Epidemiological Monitoring_

Using Forest Fire Data to Improve **Machine Learning Modeling** of Mosquito Abundance
Predicting Culex Mosquito Habitat and Breeding Patterns in Washington D.C. Using **Machine Learning Models**

Predicting Mosquito Peaks in Chicago Summer using **Remote Sensing** Climate Data and Machine Learning

West Nile Virus: Relation to Land Cover and Population Density

The Effects of Roof Design on Mosquito Activity

The **Correlation** Between Annual Precipitation and West Nile Virus in the United States

A comparative analysis of different **Convolutional Neural Network Architectures** for Mosquito Genera Classification

Predicting the Volume of Mosquito-Borne Diseases Via the A.M.E.A. **Radio Network. Land Cover Verification and Error Analysis** for Citizen Science Applications

Do mosquitoes have a desired choice of host determined by socioeconomic factors?

How does air pollution, humidity, precipitation, and temperature influence mosquito-borne vectors in the United States?

Method for Effective Mosquito Data Classification to Identify Potential Hosts of Malaria with **AI Implications**

MASC AI: A Novel Method for Effective **Mosquito Data Classification and Mapping**

Predictive Model for West Nile Virus Surveillance by Detecting Inland Water Eutrophication Using **Sentinel-2 Imagery**

A Mosquito is Worth 16x16 Larvae: Evaluation of **Deep Learning Architectures** for Mosquito Larvae Classification.



Osemeren, SEES Earth System Explorer 2023

29 Nov - 1766 Views

Adopt-a-Pixel | Where does water meet land in my community? Scanning the Area As our first assignment as part of the Earth System Explorers team, we've been tasked with making our first observations using the GLOBE Observer App. The goal was to make 12 total observations (6 Potential Mosquito Habitat & 6 Land Cover). To be more efficient, I decided to make all of my observations...



Ryan C, SEES Earth System Explorer 2023

29 Nov - 1641 Views

I built my mosquito traps on June 20th and ran the experiment until July 15th. I collected a total of 15 observations during this period. To construct the traps I used three identical pots to hold water. The goal of my experiment was to find which type of mosquito bait would be the most effective at attracting and sheltering mosquitos. For the three baits I used dried leaves, grass clippings,...

Hugo, SEES Earth System Explorer 2023

28 Nov - 1704 Views

AOI Fieldwork At first, collecting land cover data for 37 points within my area of interest felt like a daunting task. However, after only a few "expeditions" in the past week, I was able to complete all 37 observations. To make my work easier and more accurate, I intentionally choose an AOI that would contain as many points as possible within a nearby park. This was to ensure that I could...



Awards and Recognition

2022

Outstanding Student Presentation Award (OSPA) winner for the American Geophysical Union's (AGU's) Earth and Space Informatics Division, (in competition with undergrad and graduate students)

Analyzing the Correlation between Terrain and the Characteristics of Wildfires.
 Trisha Goran, Kate Bittle, Renee Easterbrook, and Dafne Rodriguez
 Institute for Global Environmental Strategies, NASA Earth Science Education Collaborative

Abstract: Wildfires, also known as bush or forest fires, pose significant threats to land areas, wildlife, and property. This study examines the correlation between terrain and wildfire characteristics in response to fire. Analysis is based on average elevation, and average slope, while also considering variables like wind speed and temperature. It then explores the terrain effects on wildfires by first characterizing wildfires in terms of surface reflectance, spread, and landscape changes, and then using that data to find correlations with terrain. The research was done using publicly available data from the GLOBE Observer and the National Earth Explorer by NASA, NOAA, Climate.gov, NCEI, NOAA, and USGS. The study found that there is a strong correlation between terrain and wildfire characteristics. Based on this study, we recommend additional variables like humidity and biomass activities to better assess the effects of terrain on wildfires.

Introduction: In 2020, the US experienced a particularly devastating wildfire season, with 80,200 wildfires reported. These fires resulted in a whopping 7,000,000 acres of land, leading to widespread destruction of natural habitats and critical ecosystems. The destruction of natural ecosystems disrupted natural processes, including the nutrient cycle, soil, and water cycle, and hindering ecological recovery. The flames and ash radiating from wildfires can also have a significant impact on human health, including the release of toxic smoke and particulate matter. The wildfire made environmental efforts increasingly difficult for firefighters, who frequently worked dry and steep, reducing the impact and preventing communication at risk. Not all wildfires are created equal, however. Based on location, wind speed, and terrain, the effects caused by wildfires can vary significantly. Understanding the way that terrain affects wildfire spread can help us better understand the way that terrain affects wildfire spread. This study aims to determine the specific terrain factors that can affect the spread and direction of a wildfire. We primarily focus on California due to the state's unique geographic, topographic, and vegetation characteristics, which play a significant role in determining fire behavior. Steep terrain can act as a barrier for fire propagation, accelerating its spread and making firefighting efforts even more difficult. In contrast, the terrain may also help control fire, allowing firefighters to establish firebreaks.

Research Question: How does terrain affect California wildfires in terms of wildfire intensity in terms of surface reflectance, land cover change, and their spread?

Analysis of Results: Wind speeds accelerate wildfire extinction by extinguishing flames and reducing temperatures, thus lowering wildfire occurrence rates. Greater wind availability shortens wildfire duration, evident in correlated graphs. Collected data establishes direct slope-angle correlation with wildfire spread, simplified by wind speed and climate factors. Mountainous woodlands, with dense tree spacing and steeper slopes, are prone to extensive wildfires. Heat-driven upward flames reach nearby vegetation, sharply dropping the Base 1000 index. Wind reinvigorates during burn side spread, surface reflectance and slope angle show equivalent ratios, with increased slope angles decreasing reflectance. Recent years saw most large-scale fires in high-elevation California areas, possibly due to dense forests and dry air facilitating rapid vegetation ignition. GLOBE Observer's table 4 data confirms significant post-fire land cover changes, validating altered reflectance and terrain effects. Observations highlight road barriers and water bodies' efficient wildfire spread prevention and vegetation change reduction.

Aligning GLOBE Observer Mosquito Habitat Mapper and Land Cover Citizen Science Datasets to Open Geospatial Consortium Standards

Prachi Ingle¹, Matteo Kimura¹, Anne Bowser², Jonathan Fisk³, Alex Long³, Russanne Low⁴ and Peder Nelson⁵, (1)University of Texas at Austin, Austin, TX, United States, (2)Woodrow Wilson Center, Washington, DC, United States, (3)Woodrow Wilson Center, Washington, United States, (4)Institute for Global Environmental Strategies, Arlington, VA, United States, (5)Oregon State University, Corvallis, OR., United States



BACKGROUND

Citizen science, where volunteers collect scientific data, is producing Earth science data at a scale that offers significant and unique contributions to scientific research. However, each citizen science project often has their own rules, procedures, and protocols which makes these datasets less interoperable, thus limiting their potential impact. Thus, the Open Geospatial Consortium (OGC) presents a standard for geospatial data to promote interoperability through shared definitions and conventions between various datasets. GLOBE Observer Mosquito Habitat Mapper (MHM) and Land Cover (LC) datasets served as the first test case for aligning citizen science data to OGC standards:

- The GLOBE Observer Program:** GLOBE is a NASA-affiliated citizen science initiative. Through a mobile app, users can collect data on trees, mosquitoes, land cover, and clouds. This data is publicly available via the GLOBE Observer API.
- Global Earth Challenge (GEC):** A joint initiative by the Woodrow Wilson International Center for Scholars (The Wilson Center), EARTHDAY.ORG, and the U.S. Department of State to encourage citizen science datasets to align with the Open Geospatial Consortium (OGC) SensorThings API interoperable data standard.
 - OGC SensorThings API:** A standard for diverse Internet of Things (IoT) sensor systems. Citizen scientists use sensor systems (i.e. phones) to conduct observations of their local areas. We worked alongside GLOBE Observer and Global Earth Challenge teams to harmonize the GLOBE Observer Mosquito Habitat Mapper and Land Cover to the OGC Interoperable data model.

METHODS

The Global Earth Challenge (GEC) provided training and a template for us to align GLOBE Observer Mosquito Habitat Mapper (MHM) and Land Cover (LC) datasets to OGC's SensorThings API. Below we enumerate the primary components of this alignment procedure:

- Curated citizen science datasets** – Converted the raw GLOBE datasets into a more intuitive and research standard for use in the next steps.
- Described Metadata**
 - Specified Results** – We described the datasets' overall purpose, identified specific subjects of study, and described which data fields were collected by humans or sensors.
 - Specified Data Quality Fields** – We explained the data quality metrics and validation procedures from the curated GLOBE Observer MHM and LC datasets.
 - Provided Attributions and Licenses** – Crediting the data aggregator empowers users to more confidently understand and use GLOBE Observer data in their own research.
- Describing Field-Mapping** – We described the purpose, units, and observation types of each field name data column. We indicated which fields prove most useful for Users. We also highlighted key information about image data collected in both the MHM and LC datasets.

RESULTS

These alignment procedures better contextualize the GLOBE Observer MHM and LC datasets. The resultant metadata is publicly available via the Citizen Science Cloud Portal.

Following our alignment, several other citizen science initiatives have also participated in the Global Earth Challenge. Examples include datasets studying mosquitoes, bees, plastic pollution, and air quality. Overall, these datasets have totaled to over 55 million data points across over 8 citizen science datasets. Current researchers are synthesizing GLOBE datasets with other OGC-aligned mosquito datasets to document mosquitoes in Florida¹.

Thus, the alignment and usage of OGC-aligned datasets highlights the potential in harmonizing datasets to the OGC SensorThings API standard.

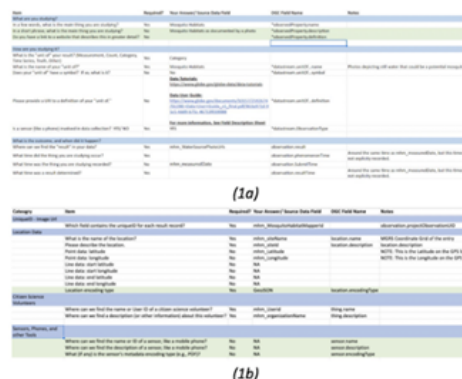


Figure 1: These tables depict key steps in the OGC Alignment procedure for the GLOBE Observer MHM dataset. (1a) depicts the primary results for this dataset. (1b) depicts overall metadata for this data.

CONCLUSIONS

This OGC alignment for GLOBE Observer MHM and LC data can serve as a template to crosswalk other specialized GLOBE datasets to the broader, more current OGC SensorThings API standard. Our work contributes to ensuring citizen science data interoperability across projects part of the Citizen Science Cloud Platform.

This demonstrates the efficacy and necessity of cross project collaboration and interoperability standards in the world of citizen science. Citizen Science interoperability will improve the accessibility and reusability of citizen science data to answer and monitor the pressing environmental questions of the future.

ACKNOWLEDGEMENTS

We would like to thank our project mentors, Dr. Russanne Low, Peder Nelson, and Andrew Clark, for their guidance throughout our project.

The material contained in this poster is based upon work supported by National Aeronautics and Space Administration (NASA) cooperative agreements NNX16AE28A to the Institute for Global Environmental Strategies (IGES) for the NASA Earth Science Education Collaborative (NESEC) and NNX16AB89A to the University of Texas Austin for the STEM Enhancement in Earth Science (SEES). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

REFERENCES

- [1] The Woodrow Wilson International Center for Scholars. (2021). *Global Earth Challenge Key Achievements, Impacts, and Opportunities*. Global Earth Challenge. <https://storymaps.arcgis.com/stories/37f9b97abde8419c-8992-d7867e4e6e92>
- [2] The Woodrow Wilson International Center for Scholars. (2021). *Maximizing the Impact of Citizen Science*. Global Earth Challenge. <https://storymaps.arcgis.com/stories/25a85168d63c4955b8e1ad7d442bf6838>
- [3] The Woodrow Wilson International Center for Scholars. (2021). *Sharing Citizen Science Data*. Global Earth Challenge. <https://storymaps.arcgis.com/stories/30b18d96dc04d3294be11654394c49>



SEES Earth System Explorers Intern Recognition

A Novel Approach for Predicting Large Wildfires Using Machine Learning Towards Environmental Justice via Environmental Remote Sensing and Atmospheric Reanalysis Data across the United States

Nikita Agrawal
Whitney M. Young Magnet High School, Chicago, IL, USA

Introduction

Large wildfires, ≥ 125 hectares burned, account for 95% of total annual burned area in the US per the USDA.

Figure 1: Total million hectares burned in the United States from 1985 to 2022.

Current Predictive Models - Localized, expensive computation

Research Questions and Goals

- Can large wildfires be predicted with high accuracy across multiple regions using easily available data?
- Which environmental and atmospheric variables are significant in impacting large wildfires and how can these variables be identified?
- How can the prediction of large wildfires be applied towards environmental justice impacts?

Figure 2: Engineering goals of research.

Wildfire Sites Analyzed

Figure 3: Plot of United States map depicting 2109 wildfire sites, retrieved from USDA, representing 1.3 billion NASA MODIS observations.

Red points represent large wildfire sites with a burn area ≥ 125 hectares. Purple points represent not large wildfire sites with a burn area < 125 hectares.

Variables Used

Derive 38 total environmental and atmospheric variables

Figure 4: Derive 18 environmental variables.

Figure 5: Derive 20 atmospheric variables.

Methodology

Input → Processing → Modeling and Tuning → Evaluation

Step 1: Input

- ERAS: Atmospheric reanalysis data - ECMWF
- NASA MODIS: 1.3 billion MODIS Observations - Oak Ridge National Lab
- COUNCIL ON ENVIRONMENTAL QUALITY: Justice 40 Initiative
- USDA: 2109 wildfire sites 14 million hectares burned

Figure 6: Input sources.

Step 2: Processing

Figure 7: 10 km by 10 km grid for which environmental data was sampled from.

Step 3: Modeling and Tuning

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost
- KNN
- SVM

Figure 8: Using these six selected machine learning classification models, implemented k-fold cross validation process with k value of 10.

Step 4: Evaluation

- Two Model Validation Techniques
- Model Accuracy Analysis 6 Models
- Variable Importance 38 Variables
- Environmental Justice Mapping

Environmental Justice

Fire safety organizations can leverage the model created to:

- Predict potential large wildfires with greater accuracy. Can be used to employ preventive safeguards early on and optimize resources to reduce the spread of wildfires.
- Identify vulnerable disadvantaged geographical areas that will also be impacted by wildfires, to act with high priority when allocating federal assistance.

Figure 12: Disadvantaged communities that are also impacted by wildfires.

Red points represent large wildfires with a burn area greater than or equal to 125 hectares from 2018-2020. Purple points represent non-large wildfires with a burn area greater than or equal to 125 hectares from 2018-2020. Orange points represent disadvantaged communities per White House Council of Environmental Quality Justice40.

Conclusion

Analyzed 2109 wildfire occurrences, which have burned 14 million hectares, across the United States from 2000 to 2020.

Improvement from existing predictive models

XGBoost Classification model performed the best in predicting large wildfires with an accuracy of 90.44% and had the greatest AUC of 0.97.

Figure 13: ROC Curve comparison of all 6 models developed in this research.

Model Type	Accuracy score
Logistic Regression	69.81%
Decision Tree Classification	80.19%
Random Forest Classification	87.62%
XGBoost Classification	90.44%
KNN Classification	67.48%
SVM Classification	69.95%

Table 1: Accuracy score of the machine learning models used in this research.

Environmental Justice impact when applying the XGBoost model towards identifying disadvantaged communities at risk of large wildfires, to prioritize resources.

Limitations and Future Research

To improve our understanding of the effect of wildfires, future research could incorporate data on human impacts. Optimized routing paths can be modeled to help send resources expeditiously to vulnerable disadvantaged communities impacted by large wildfires.

Figure 14: Future research could include data on direct human impacts, such as those that cause wildfires through ignition, suppression, or altering fuel distribution.

2023

Nikita Agrawal: Research on Predicting Large Wildfires Using Machine Learning

- **Regeneron Science Talent Search (STS) Scholar** - Society for Science
- **Third Place Grand Award**, National Geographic Society, Earth and Environmental Sciences Category
- **1st Place in Earth Science Category**, Illinois Junior Academy of Science
- **National Geographic Society: Excellence in Geography and Geospatial Science**
- **Plus 8 additional awards** for this student research project.

2024

- Three SEES ESE intern projects were in the Regeneron top 300 (in 2023, one was in the top 6)

Research publications including SEES Interns (9)

2023 Agrawal, N., Nelson, P.V. and Low, R.D., 2023. A Novel Approach for Predicting Large Wildfires Using Machine Learning towards Environmental Justice via Environmental Remote Sensing and Atmospheric Reanalysis Data across the United States. *Remote Sensing*, 15(23), p.5501.(SEES 2022)

2022 Sun, C., Nimbalkar, J., & Bedi, R. (2022, September). Predicting Future Mosquito Larval Habitats Using Time Series Climate Forecasting and Deep Learning. In *2022 IEEE MIT Undergraduate Research Technology Conference (URTC)* (pp. 1-5). IEEE. (SEES 2022)

2022 Xia, I., Singirikonda, N., Hellman, L., Watson, J., Hanna, M. and Low, R., 2022. Using Machine Learning Regressors for the Discovery of Culex Mosquito Habitats and Breeding Patterns in Washington DC. *Journal of Student Research*, 11(4). (SEES 2022)

2022 Polineni, S., Shastri, O., Bagchi, A., Gnanakumar, G., Rasamsetti, S., & Sundaravadivel, P. MOSQUITO EDGE: An Edge-Intelligent Real-Time Mosquito Threat Prediction Using an IoT-Enabled Hardware System. *Sensors*, 22(2), 695. (SEES 2021)

2022 Low, R.D., Schwerin, T.G., Boger, R.A., Soeffing, C., Nelson, P.V., Bartlett, D., Ingle, P., Kimura, M. and Clark, A. (2022). Building International Capacity for Citizen Scientist Engagement in Mosquito Surveillance and Mitigation: The GLOBE Program's GLOBE Observer Mosquito Habitat Mapper. *Insects*, 13(7), p.624. <https://doi.org/10.3390/insects13070624>. (SEES 2020)

2022 Low, R. D., Nelson, P.V., Soeffing, C., Clark, A., and SEES 2020 Mosquito Mappers Summer Research Interns. Adopt a Pixel 3 km: A multiscale data set linking remotely sensed land cover imagery with field-based citizen science observations. In *Open Citizen Science Data and Methods*, ed. Bowser, A.K. Schaed, S. and de Serbinin, A., *Frontiers Research Topics* <https://www.frontiersin.org/research-topics/13843/open-citizen-science-data-and-methods>.

2021 Low, R., Boger, R., Nelson, P. and Kimura, M. GLOBE Observer Mosquito Habitat Mapper citizen science data 2017-2020. Published August 19, GeoHealth. DOI: 10.1029/2021GH000436. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GH000436>. (SEES 2020)

2021 Kar, K. and Low, R. The effect of poverty on mosquito-borne illness across the United States. *J Emerg Investig.* February 25. <https://www.emerginginvestigators.org/articles/the-effect-of-poverty-on-mosquito-borne-illness-across-the-united-states>. (SEES 2020)

2020 Lin, D.W. Land Cover Feature Analysis for Mosquito Habitats. *IOSR Journal of Environmental Science, Toxicology and Food Technology*. e-ISSN: 2319-2402, p- ISSN: 2319-2399. Volume 14, Issue 12 Ser. II (December 2020), p 24-30. https://www.academia.edu/44798039/Land_Cover_Feature_Analysis_for_Mosquito_Habitats?email_work_card=thumbnail (SEES 2020)



Cassie Soeffing

Cassie_Soeffing@strategies.org

