Forest Carbon Codefest: Promoting discovery in environmental data science through a collaborative coding event

Tyler L. McIntosh, Nayani T. Ilangakoon, Katie J. Ellis, Megan E. Cattau, Megan K. Littrell, Ty Tuff, Erick Verleye, Carol A. Wessman, Jennifer K. Balch

Event Overview

BY THE NUMBERS

2.5 days
28 participants
6 teams
3 training events
>50% participants from industry

The Forest Carbon Codefest, hosted by Earth Lab and the Environmental Data Science Innovation & Inclusion Lab (ESIIL), took place from March 12-14th, 2024. At this event, six teams of 4-5 scientists, land managers, and industry partners used open cyberinfrastructure and data to define their own scientific questions that advanced understanding of aboveground forest carbon dynamics in the Southern Rocky Mountains. Teams then collaboratively answered that question in the form of a fully open, reproducible, and complete code script.

Goals of the event included promoting innovation, inclusion, and open science; establishing new collaborations, and providing space and structure for playful data exploration and analysis.

Balance between four essential pillars enabled innovation in a fast-paced hackathon-style event for environmental data scientists

Diverse participant community

The event brought together a diverse community from across multiple scientific disciplines and career stages, with 54% of participants from community partner organizations such as private corporations, government agencies, and nonprofits.

Although varied perspectives, experience levels, interests, and coding languages were all mentioned as particularly challenging aspects of the event, significant percentages of participants surveyed felt better prepared for cross-disciplinary innovation following the event.

I am better prepared for community-building and knowledge exchange

83%

I created new connections across disciplines and sectors

95%

Refined yet flexible scientific mission

The Forest Carbon Codefest was unique in that it was a hackathon-style event, but without a single specific task for the participants. Rather, teams were asked to design their own question within a set of subject area and geographic constraints.

In a post-event survey, participants highlighted time constraints and the significant volumes of data as overwhelming, mirroring many of the everyday realities in the burgeoning field of environmental data science. Throughout the event, it was apparent that the constraints set by the facilitation team on project topic and scope were essential for allowing teams to exist in a space of creativity while also coalescing quickly around a single actionable idea.

"Given the time constraints and volume of data, it was hard to think really outside the box. But some groups came up with some very creative ideas in very little time, which was cool to see!"
~ Anonymous participant

Accelerated data access

The event's curated data library was composed of pre-downloaded data in addition to data access scripts for mounting commonly used datasets for forest carbon research. Tutorials were provided to participants for accessing additional data via STAC and VSI that was beyond the scope of the data library.

When asked about the curated data and access tools, participants said that the data library was "inspiring," "comprehensive," "useful," and "facilitated effective use" by the participants. Times when data access broke down were emphasized by participants as frustrating, highlighting the importance of streamlined data use in a time-constrained environment. Numerous participants also mentioned being exposed to new datasets and resources as a key outcome of the event.

Scan the QR code at the top of this poster to check out the compiled data library and other resources from the event!

Cloud collaboration training & resources

Cloud-based collaboration tools and resources were the connective tissue necessary for a diverse set of participants to work together on a data-intensive, mission-oriented project, reflecting the reality of our rapidly evolving scientific world.

Surveyed participants said that working collaboratively in real-time on cyberinfrastructure and platforms like GitHub was both challenging and rewarding, providing hands-on experience in group coding and project workflows. The importance of training in cloud collaboration tools and platforms prior to the event was highlighted by cited challenges in tandem with significant improvements in participant confidence before and after the event in related skills.

The training on skills for cloud-based collaboration prepared me for the event

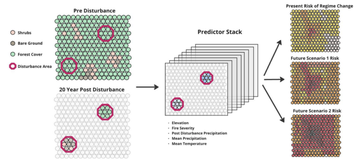
89%

Event Outcomes

In only about 10 hours of dedicated working time, teams at the event were able to select scientific questions of interest, orient to available data, and begin analysis. All teams were able to generate multiple meaningful visualizations or quantitative analyses, and documented their code in an open GitHub repository accessible via the event website (see QR code).

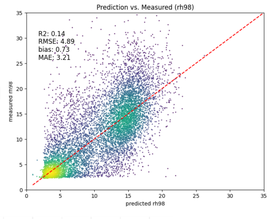
Team project topic areas ranged from generating a wall-to-wall canopy height model from Landsat and GEDI data to analyzing post-disturbance trajectories of aboveground biomass across the Southern Rocky Mountain Ecoregion, and even spanned to questions like, "Where are sawmills, can we identify forest products, and do they impact the surrounding forest?"

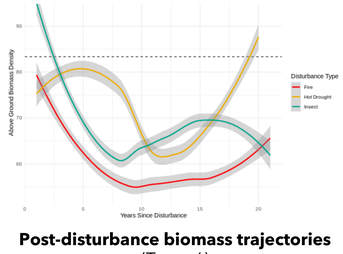
The below figures were generated by teams during the event, and serve as an examples of some of the types of projects that were in development.



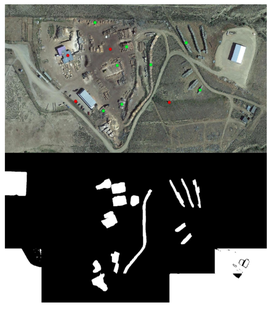
Model visualization for predicting ecosystem transformations (Team 1)

Canopy height model performance (Team 2)





Post-disturbance biomass trajectories (Team 6)



Sawmill product classification (Team 4)

SURVEY RESULTS

Event served as an incubator of ideas

94%

"To a 'Moderate,' 'large,' and 'very large' extent"

Overall satisfaction

78%

Meeting was well planned & designed

100%