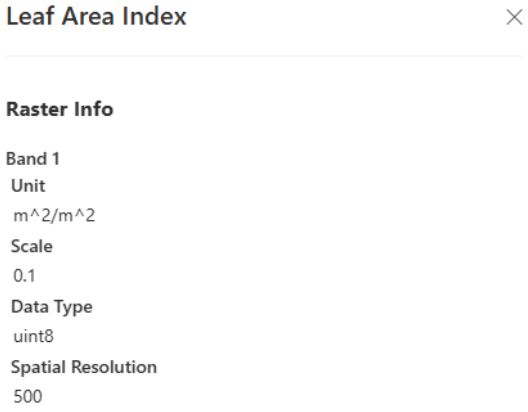
June 30, 2025 Joint Methods & CI Working Group Call

Attendees: Eil Horner, Dave Durden, Jody Peters, Brittany Barker, Chris Jones, Lindsay Campbell, Josie Hughes, Will Hammond, Hassan Moustahfid

1. Spatial Forecast Workshop for the EFI2025 Conference recap and debrief
    a. Workshop tutorial material is at: https://github.com/eco4cast/modis-lai-forecast/tree/main/tutorials
    b. The rendered html is at: https://htmlpreview.github.io/?https://github.com/eco4cast/modis-lai-forecast/blob/main/tutorials/efi_2025_workshop.html
    c. Need to update workflows to multiply LAI by its scaling factor
        i. 

    d. Debrief from conference
        i. There were 2-3 people applying spatio-temporal forecasting challenge towards phytoplankton blooms - Caroline Owens is interested in this
            1. Connect Hassan with Caroline since the HABS was something of interest that came up in a discussion about marine/Great Lakes forecasts
        ii. Most people who attended the workshop were not interested in the fire aspect of the challenge, but were more interested in creating spatio-temporal models
        iii. All the models used could be ported over to another system
        iv. Climatology has been the null model to beat.
        v. The big thing to add is a simple model to show how to do a simple spatial-temporal interactions
        vi. Everything happens at each cell, but there is not spatial covariance. You might as well forecast each cell individually
        vii. Another thing to do is to change the resolution - there are averaging parameters in the data grabbing functions that may have been better to do, but there is a balance between amount of data vs averaging. To keep it cheaper for forecasting, needed to keep resolution broad so we didn't run out of the bucket space

1. Tiffs can get really big
2. Scale is the big issue to consider. Second consideration is comparing across space

viii. If climatology had been restricted to after the fire or similar type of fire events in other ecosystems - could be useful, but would need to set up

ix. One thought would be to switch to Google Earth Engine to avoid needing to hold the data in the bucket. The Planetary Computer was working only part of the time.

1. Provide code to make it easy for people to get what they want without needing to host it
2. This seems to be where the people building bird models in Canada are moving to - sometimes it takes a long time, but it does make it more doable to do things
    a. Josie can share the code that has been done with the birds - send her an email to

e. For the workshop example - used one fire and then showed out to expand it out.
    i. But Dave did test other fires while setting up the workflow
    ii. Great Smokey was pretty small - would have come out to 2 cells
        1. So if we go down to a native resolution, this would have been better
    iii. Focused on bigger fires for the workshop
    iv. The Creek fire at the Santa Rita site was a good size for the challenge
f. What about pulling tiffs and turning them into flat files? Then have predictions shared out as rasters
    i. Didn't do this for the workshop
    ii. Brittany converts rasters into a matrix form that saves space
    iii. Think you might use some functionality moving to geotiff in cloud based system and being able to use STAC API
g. Any interest from Will, Chris to do a seminar/record the tutorial?
    i. Will has some bandwidth
    ii. Would be nice to spend some more time on the cyberinfrastructure since there seemed to be lots of interest in using the infrastructure to apply to other domains
    iii. Nice to break it up into a smaller series - one on just infrastructure and then others on the challenge
h. For the conference - the group modified the forecasting infrastructure to be able to demonstrate the scoring but that meant doing more hindcasts since the new observations were not available to do the scoring
    i. A task for Will is to revert the workflow to set it back to forecast now that the conference is over.
    ii. When people submit the forecast you can't score them right away. Have to have a queue to wait for the new target observations to come in
i. Issues of pulling data efficiently and thinking of scale and decisions need to be made

        i. Allow modelers to determine what scale to consider with the decisions they were making
        ii. Think about allowing modelers to decide what scale to use.
        iii. People setting up forecast challenge will think about scale for scoring.
        iv. Score at a subset of pixels that are small and then others can decide what scale they want to use

2. What are people most interested in?
    a. Brittany - has done ecological modeling of fires and the cyberinfrastructure - curious at how forecasts perform at different sites. If it doesn't perform as well
        i. And figure out what decision makers want to know
    b. Lindsay - cyberinfrastructure workflow and how it can be generalized to multiple problems - happy to contribute to any ecological process
        i. Modeling people need to make decisions and how will the forecasts be used
        ii. What are the workflows and how can they be accessible to a broader range of users (modelers)
        iii. What is the plumbing - what can people use as tools to apply to their own systems
    c. Hassan -
        i. Scalability and Performance, Bridging the Research-to-Operations Gap. CI- Security and Reliability, Standard Best Practices and Automation of Iterative Forecasting Workflows
        ii. Uncertainty Quantification, I think we addressed this at some level.
        iii. Data Accessibility and Sharing
        iv. Data Volume and Velocity (Big Data)
    d. Eli - joining to listen and learn
    e. Josie - How to connect with community to help with big tasks working on a shoe string budget with minimal people. Cross fertilization of ideas and tools

3. Note from May 2025 call and the follow-up Slack discussion:
    a. Blog post ideas that came up on Slack - how to set up a forecasting CI workflow and then how to expand when you outgrow your initial workflow (see the Slack discussion below)
        i. Would need Carl, Ethan, Mike, Quinn to lead such a post
    b. How to get the forecasting challenge workflow accessible to others without always needing Carl/Quinn to help with the secret sauce
        i. Hard part is the CI has to be tied to someone at EFI to get credentials and get the buckets set up
        ii. There could be a way to set up a folder that is shared with just a few people and the secrets could be shared on an as needed basis.

c. Unconference topic or multiple day workshop - applying the forecasting pipeline to develop your own challenge. Think this would be really useful for the community
d. John reached out to Eliot McIntyre (Canadian Chapter) about scoring spatially misaligned forecasts. Your model does really well but your predictions are a pixel off. Eliot has a manuscript he will be posting soon about spatial misalignment
e. Looking for advice or technical documentation about creating a new forecasting challenge - for now, advice is to fork the neon4cast-ci GitHub and then compare the neon4cast-ci and the usgscast-ci to see what needs to be tweaked, but it would be nice to have a technical document to walk through the steps.
    i. Jake Zwart has rough notes about the repository setup here https://github.com/eco4cast/usgsrc4cast-ci/blob/main/repo_setup_notes.txt
    The main difference for getting started is in Jake's initial PR https://github.com/eco4cast/usgsrc4cast-ci/pull/1
f. It would be helpful to have a schematic/documentation about which keys/secrets are used for what, how/where the data is stored, and the relations/hierarchy between buckets/containers/actions/STAC files?
g. Would be good to also document modularity (which parts are dependent on each other or not), and functionality (which parts are static/dynamic)