October 27, 2023 Joint Methods & CI Working Group Call

Attendees: Quinn Thomas, John Smith, Jody Peters, Carl Boettiger, Dave Durden, Emma Mendelsohn, Chris Jones, Libby Mohr, Brittany Barker

Agenda/Notes:

1. Any workshops, panels, short courses, etc to propose for the EFI 2024 Conference?
    a. Due: November 1. Use this short feedback form to submit proposals for workshops, panels, short courses, socials before/after/during this conference.

2. CI/Methods needed for spatially explicit forecasts: Forecasting Wildfire Recovery Using MODIS Leaf Area Index
    a. GitHub repo: https://github.com/eco4cast/modis-lai-forecast/
    b. TERN example to use as reference: https://projects.ecoforecast.org/tern4cast/
    c. Tasks to set up GitHub Action Workflow: https://github.com/eco4cast/modis-lai-forecast/issues/10
    d. Check in with John about the random walk model
        i. John has the model, tested it and it works. Now working on getting the raw tiff files into the correct format to pass into NIMBLE. Having some software issues. Working with Carl to troubleshoot
        ii. Question about formatting - want CI for the spatial tiff file to be parallel what is currently used with csv/netcdf
            1. For example, want to have expected columns in parquet in tiff
            2. How are model ID, reference date time being captured?
                a. Want to be partitioning the directory structure. The team is not a single file, it is a collection of tiffs. So want to think about how to set this up
                b. Have a spatial bucket that the files get submitted to.
            3. Is the info in the file name or in the tiff?
            4. In the normal Challenge csvs/netcdf get sent to forecasts submitted to a bucket. Then that gets checked that the forecast is in the correct standard and then it is modified into a parquet file that then goes to the forecast bucket to then be scored and visualized
            5. Need to clean up the workflow so that there is a script that regularly runs to targets and regularly processes the submissions
            6. With the standard Challenge the NEON forecasts go to buckets for each theme and then in each theme the models go in as a parquet files that have the model names, etc
            7. Quinn/Carl working on options for partitioning structure to have maximum flexibility across data types (NEON data streams, USGS data stream, TERN data stream) and duration (half hourly, hourly, daily, monthly - this is the averaging period)
            8. Duration discussion

a. LAI mean for a duration would be the LAI mean for that duration. But could do LAI max or LAI mean which would be the max or mean of that duration
b. Have the bucket folder set up by variable name, then aggregate - because you can aggregate variables different ways
c. For the MODIS LAI it is currently the average from the month
d. Different Challenges could match up durations to compare across variables from different Challenges
e. If we put tiffs in a parquet function? Will it read them?
    i. Better to have a folder for only tiffs that come from a particular model
    ii. Forecast folder -> tiff -> then follow the same structure used in the standard Challenge
f. Duration is not in the current EFI Standard, but it is in the STAC and follows ISO. In the Standard it is sort of represented by the start time/end time, but need to discuss this further

e. Land cover information to use for making predictions https://eros.usgs.gov/doi-remote-sensing-activities/2020/usgs/rangeland-fractional-components-across-western-united-states
f. Opportunities for engagement with the working group
    i. Hands-on: Interest in setting up co-working times to work hands-on with the code?
    ii. Big picture: What fire polygons do we want in the Challenge? Want them to cover a range of vegetation types/ecoregions and range of ages. Will need good documentation and metadata
    iii. Brittany put Carl/Jody in touch with Justin Welty from USGS. Justin manages a geodatabase of wildfires and the Wildfire Fire Trends Tool. Justin is willing to meet. Trying to see if it will work best to have Justin join a call or to schedule a call with a subset of the group
g. Next steps
    i. John is continuing to work on random walks model
    ii. Clean up code for the workflow
    iii. Parametric Scoring. The group has completed the the non-parametric scoring.
        1. John has a student who has experience with the scoring rules package. Think he can make it work for this for the parametric scoring
    iv. What goes into the scoring function?
        1. This will scale terribly if you have more pixels compared to the ram you have

2. Think what is set up right now will work for now, but will eventually want to set something up to work with larger datasets
3. Chris can work on this after Dec 8
4. Will need to deal with NAs
    a. Currently using the na_bottstrap_fun function - if doing this backend in terra will need an analogous routine

v. Dave - interested in working on the GitHub actions, particularly working towards target generation

vi. Target generation - start with the target. Take the firebox code out from the quarto and the MODIS data and create the targets

1. The AWS_ACCESS_Ketys in the GitHub actions is what allows it to run

vii. Add everyone as direct contributors to the repo

viii. Can speed up code in places

ix. Emma - can work on the tiff submission and validation checks

x. What does this look like with multiple polygons?

1. What if you have 2 sites?
    a. Do you put it in 2 folders? The spatial data is attached to the tiffs, but you have to look in each one to see where it is covering
    b. So should this be part of the folder structure or should it be in the file name
    c. Decided to give each polygon a siteID
    d. Eventually could make STAC catalog over the object and user could query to get just the tiffs in the polygon of interest
2. Would like to add at least one more polygon
    a. If people have a site that greens up faster that would be interested
3. This does not have to be only for forested ecosystems. Could be for all ecosystems.

xi. Can change this out from MODIS and use NDVI

1. The current example does not have much resolution. Do we want to increase resolution spatially and temporally

xii. USGS has all of the US fire products polygons going back to late 1800s

1. Resources from Brittany: geodatabase of wildfires that occurred in the western US (maybe just the Great Basin?) over the past several decades. It looks like they published a "Wildfire Fire Trends Tool" in 2022 that uses these data.This could be useful for development or validation of your fire forecasts.
2. The other group I mentioned is the USGS EROS group over in Sioux Falls, SD. They have developed a Landsat-derived landcover component dataset for the Great Basin consisting of fractional estimates of functional groups such as shrub,

sagebrush, perennial herb, and annual herb. I published a paper that integrated this dataset with field monitoring datasets to explore plant community changes over time in response to fire.

xiii. What does recovery mean for this forecast?
    1. We are not defining it, it is LAI response, but doesn't get at what species are coming back.
    2. Need a data product that provides the data that is of most interest
    3. Can estimate LAI from Sentinel and Landsat. Mike has module in PEcAn
        a. If want to stay with LAI at a finer spatial resolution could work on using this

xiv. Can use this as spring board for a proposal for when SVG comes out on post fire recovery.

xv. People can claim GitHub issues - https://github.com/eco4cast/modis-lai-forecast/issues/10
    1. Select an issue from the list in the link above and save it as its own issue
    2. Carl is happy to help get people set up as they work through GitHub Action questions

xvi. Visualization? Do we want them to be interactive?
    1. First priority - score the forecast
    2. If we can say here are forecasts of LAI that are made every month and this forecast is better than another one.
    3. With the spatial forecast - each location could be accurate in one location but horrible in another
    4. Once we have the forecast up then we can start asking questions about how well the forecasts are doing across sites

xvii. Could have a subgroup that works on identifying polygon, what is the data product that is measuring what is useful for making forecasts and is ecologically relevant
    1. Once the structure is set up then we can swap out data products
    2. Doug Shinamin also at USGS where Justin works

3. Other Updates
    a. CI Workshop Proposal Update (Jake, Jessica, Chris);
        i. Goal: Identifying and filling gaps in CI/Methods for forecasting. Bringing together people from gov't agencies, academia, and private sector/NGO
        ii. Due to limited space, invites went out to ~50 people on Wednesday!
    b. Standards Manuscript Update - Mike submitted revisions to the page proofs so it should be coming out soon
    c.