

September 29, 2023 Joint Methods & CI Working Group Call

Attendees: Alex Chubaty, Dave Durden, John Smith, Chris Jones, Carl Boettiger, Libby Mohr, Emma Mendelsohn, Jessica Burnett, Jody Peters, Quinn Thomas, Brittany Barker, Mike Dietze
 Regrets: Jake Zwart

Agenda/Notes:

1. EFI 2024 Conference Announcement
 - a. Use this short feedback form to submit proposals for workshops, panels, short courses, socials before/after/during this conference.
 - b. Deadline for proposals : 01 Nov 2023
 - c. Registration and Abstract submissions open: 01 Dec 2023
 - d. Accepted contributions are announced: 01 Mar 2024
 - e. Final program announced: 01 Apr 2024

2. Spatially Explicit Forecasts - overview and next steps
 - a. GitHub repo: <https://github.com/eco4cast/modis-lai-forecast/>
 - b. Excited to unify CI and concepts to expand a different type of forecast challenge that leverages the CI we have with the NEON Challenge
 - c. Current forecasts in Challenge are only temporal, not spatial
 - d. So moving to spatial decided at the Unconference to start with fire disturbance and forecasting the greenness of the pixels monthly after the fire occurs
 - e. Did a proof of principle that you could create the forecast in the EFI design
 - f. Read MODIS data to get the target data. Have historical data back to 2002 and get monthly MODIS average.
 - g. Created a basic forecast - ensemble average of all historical forecasts. Combined times not burned with times burned so isn't a model of recovery, but gave a benchmark
 - h. Use rasters - get geotiff from MODIS and great geotiff out that
 - i. Then able to score the forecasts
 - j. Scores how well the pixels forecasts are (small scores are better than high values)
 - k. How do you envision things being stored or organized on a bucket and maintained by EFI?
 - i. See diagram in the ReadMe in the repo
 - ii. For the file structure relative to the Challenge - Instead of having single parkay, have a set of tiffs.
 - iii. There are 3 buckets - targets, forecasts, scores that have tiffs instead of parkays
 - l. Challenge - how to scale up, how to choose spatial resolution?
 - i. Using 500m, 16 day MODIS. Coarse grained it to 1km grid cells and pooled up to 30 days.
 1. Sometimes the 30 days is helpful for dealing with missing data
 - ii. If forecasting larger area, may want to start with coarser grain.

- m. Scoring for time series in Challenge - have a column that defines family and another column that defines parameter within family. For example for normal distribution: need 2 rows - have a row for family normal, parameter mu and then have a row family normal, parameter sigma.
 - i. Scoring for MODIS is similar, just has a bit of preprocessing
 - ii. Next step for spatial project - add example of parametric scoring
- n. The code is pushing to S3 buckets, so you don't have to upload/download data
- o. John is working on implementation of random walk so we have that in addition to the climatological null
 - i. Do we want for a parametric forecast to expand to the framework Quinn has, do we read in the parameter as a tiff? Yes - have a mean tiff and a variance tiff. Would get the corresponding raster cell in the tiff
- p. In the workflow there is a GitHub actions that has been running once a month
- q. Currently, don't have a dashboard, it is just rendering the grid cells.
 - i. The quarto could be set up to run an action for targets, an action for fitting scores, and an action for visualizing forecasts.
 - ii. How are forecasts submitted?
 - 1. Tiffs are dumped into an S3 submission bucket and the Actions go and discover it
- r. Actions example from TERN (and then think about what that would mean in this example)
 - i. TERN example of a forecast in a box - it runs off GitHub actions: <https://projects.ecoforecast.org/tern4cast/>
 - ii. Go into workflows - there is a set of yamls that are the tasks
 - iii. Targets need to be autogenerated (have this for the MODIS forecast)
 - iv. People submit forecasts (including the baselines). The baselines run and get submitted to a small bucket that can receive anonymous uploads. Then have a job to process the uploads and moves them to S3 storage so people can look at them, but need credentials to put anything there
 - v. Then visualize the forecasts and submit info to STAC catalog
 - 1. The STAC catalog makes a lot of sense for the MODIS forecast
 - vi. Need to write a submission script - this takes a set of files and runs 2 things
 - 1. Validator that makes sure something isn't uploaded that breaks the scoring. It checks to make sure that the elements needed for scoring is going to be submitted
 - 2. Makes sure it was moved to where it needs to go
 - vii. Then need a process submission script
 - 1. This takes what is submitted to public bucket, check that it is valid, and then moves it to where it needs to go. If it not valid, it goes to another bucket so you know it was submitted, but it wasn't valid
- s. What is currently running is the forecast for one fire. It would be interesting to add additional fires - have collection of polygons that are of different size and different types of forests. Fires in old growth CA is going to be different than fires

in AZ. This question about the polygons to use would be a science question that doesn't exist in the other Challenge themes. Want to let forecasters wrestle with these different questions - different scales, management, age of forests, etc

- t. Have found that Challenges are successful when we have champions for the theme. People who can present workshops at conferences and can identify key groups and reach out to them.
 - i. Forest Rangeland and Ecosystem Science Center is a partner that we should connect with
 - ii. Other thing that helps people get going is the realistic example
 1. What are the things people really want. For example, people were not using the NOAA forecasts because it was hard to process the forecasts to use. We were able to make it easy to get the NOAA forecasts to help with this.
 2. So if there are similar situations for thinking about disturbance, then want to know about it
- u. Has the group looked at the land cover component datasets?
 - i. Have this for every year and could be important predictors for wildfire risk
 - ii. Having land cover and weather climate
 - iii. Brittany will share publications about the land cover in the #cyberinfrastructure Slack channel
 - iv. <https://eros.usgs.gov/doi-remote-sensing-activities/2020/usgs/rangeland-fractional-components-across-western-united-states>
- v. Alex works on fire forecasts - there are a lot of predictors that go into models including species specific/stand level information. Is there existing infrastructure to ingest arbitrary datasets? Or are the forecasts limited to specific datasets?
 - i. Open to any dataset.
 - ii. However, while we can say people can use any additional dataset, it is hard, so it is helpful if we know what kind of datasets are most helpful in order to make them more ready to feed into models
 - iii. Providing curated datasets can be tough because people may use them differently.
 1. Want to materialize those differences
 - iv. There are a few things that everyone wants and then there is a long tail of what people will want. So want to figure out what to provide and what people are responsible for calibrating their own models
 - v. Focus on what are the top things that people will want?
 - vi. Where to put issues and things to work on?
 1. GitHub Issues is best place to put ideas/comments. Slack is temporary
 - vii. Want to create this with people who are using the models so we know what is actually needed.
 - viii. Currently predicting one month ahead. Could play with the horizon - could predict one year ahead.
 1. Need to work with people who are making the predictions.

2. For example forecast, could have day after fire with mean/stdev
 3. Simple models will have the same output for all the cells in a study area
- ix. Using LAI which is a proxy for burn severity. But if there is a better proxy to use for severity, then could switch to that
 1. Could use pre-burn LAI and then use a Michaelis-Menten - what is the half saturation to get back to pre-burn value
 - x. There are MODIS optical and thermal based products
 - xi. There are lots of possible datasets, but want to get to the point of what is the model and what is the data that is needed for that model.
 - xii. Targets need to provide the full time series for the polygon pre- and post-disturbance. If internally we were to calibrate something to get back to pre-burn, we wouldn't do very well. So would need more time for recovery or need other polygons
 - xiii. Brittany has worked with the rangeland data - think they are 1x1km. Can average or aggregate as needed.
 1. Brittany uses the data to look at land use history. Have state changes. Could be important for forecasting to think about the landuse legacy of the sites. Could use the information Brittany shared here to do that:
<https://eros.usgs.gov/doi-remote-sensing-activities/2020/usgs/rangeland-fractional-components-across-western-united-states>
 - xiv. Would be good to do survey to see what would be useful for people using these data
 - xv. Use GitHub Issues to get to the point of the Challenge workflow
 - xvi. Look at existing CI for TERN project to see what needs to be done.
 - xvii. Geoparkay can be used - it is established enough to be able to use. Carl thinks it is fantastic - only will work for polygons, won't work for rasters

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
- b. Standards Manuscript Update - Mike submitted revisions to the page proofs so it should be coming out soon
- c.