

November 11, 2022 Joint Methods & CI Working Group Call

Attendees: Brittany Barker, Carl Boettiger, Quinn Thomas, Jody Peters

Agenda/Notes:

1. Data Ingest, Cleaning, and Management Task View
 - a. Didn't work on this, waiting to hear back from Matt about his availability to lead this

2. Forecasting Standards and [Challenge CI](#) Update
 - a. Quinn tested out the standards using Falling Creek Reservoir data/NOAA met and it worked
 - b. Forecast a gridded dataset hasn't been incorporated yet. It is essentially a bunch of different sites and think about how to link those up, but the datasets get really big fast, because ensemble members are needed for uncertainty
 - c. Brittany works with large spatial dataset - uses a server with multiple cores to parallelize
 - d. What is the to do list now?
 - i. Talk to more people. Have been starting with the CI on the backend for the forecasting Challenge (which is site-wise) - figure out how to march through and score that as people submit forecast. Score at scale.
 1. None of this is visible or accessible. Want to be able to allow people to access the scores and be able to create visualizations and analyses.
 2. Slowly getting closer to the front-end to what user code that people would interact with
 3. Have helper functions - but people need to know where to get the package and the helper function
 4. Have different forecasts of different variables currently have a variable column. But people. Might expect to have different columns for variable (vs having a column for air temp and oxygen)
 5. GitHub actions to automate forecast submissions
 6. There are still barriers to entry that we are working on
 - ii. Mechanically, what would need to be done to do the schema stuff to crawl. The Challenge uses a parquet dataset.
 1. Carl thinks using [STAC](#) is the way to go for this
 2. Brittany has a website that shows little images for people to click on and zoom in or be able to download rasters
 - a. Is it possible to have images/visuals to help people see what to look at instead of just listing the csv or data table, etc
 - i. Yes, think it involves brainstorming what visuals should look like.
 - ii. Will be good to tackle this

3. If building official collection of the forecast challenge - is it the csvs submitted or the parkays compiled
 - a. The STAC model make sense for where there are individual assets. Parkay is post individual assets - you can treat the whole database. So wouldn't need STAC. If we were submitted csv then would use STAC. If we point to a random parquet slice, then it wouldn't matter
 - b. Would this provide things that you need to know - what variables are being forecasted?
 - i. Yes think that would all be sucked up into the JSON metadata file
 - ii. Once we have STAC, then could button press to get schema
4. What are next steps for proof of concept to create the JSON for STAC?
 - a. Have to look at the JSON structure and look at each forecast and have to create the STAC info from the forecasts
 - b. Could group by team and reference_date_time and siteID to create an asset.
 - i. Yes, we can create an asset in any way we want.
 - ii. Could get down to a single row is an asset
 - c. STAC is a front end to discover assets
 - d. If we archive the assets on EDI, then the STAC catalog will point to where those are
 - e. Assets in STAC are typically individual tiffs, but you can choose to use csvs
 - f. Parkay takes longer to write, but it is a lot quicker to read. You can just read in a single line from a database
- iii. Other CI things to work on for the Challenge
 1. Update the code for the terrestrial 30 minute forecasts. Think that will take care of the gumming up that is happening now
 2. Quinn has been asked, why don't we forecast MODIS LAI, you can lay it on top of NEON Flux towers and people could think about how to stack forecasts. Could use an LAI forecast for input for another forecast.
 - a. This would require thinking about how to set up the targets to handle that
 - b. A tower would correspond to a single grid cell in MODIS
 - c. Could do whatever the MODIS bounding box so someone who wants to link to the Flux tower could just grab the flux they want
 - d. Are there any remote sensing products that are on a scale that you care about?

- i. Think we need to get away from NEON. Think the AOP (airborne observatory platform)
- ii. Could forecast the carbon stocks at NEON sites.
 - 1. The data latency takes awhile, so don't get feedback for awhile
 - 2. Could forecast the remote product. For example, could forecast the Sentinel satellite image
 - 3. But biomass doesn't change very much at that scale
- iii. NDVI is a good example,
- iv. LAI is also a good example (although it is very derived) because it is very dynamic and is linked to the phenology and terrestrial themes
- e. Has the group thought about Soil Moisture Active Passive (smap) Mission - Brittany is starting to work on plant species (had worked on insects and fungal pathogen) for weeds will be looking at soil moisture to see if it improves predictions for plant species
 - i. It isn't real time, perhaps 2 days latency
 - ii. For biomass the conversion of diameter to biomass comes from a model, so you are forecasting a model. Everything is a model - sensor data is a model too, but some models are closer to the raw data than others

3. Workshop Proposal

- a. Google doc to collectively develop ideas

4. For the working group - find out where people's interests are

- a. Could be teaching people the skills to vs making it easier so it isn't overwhelming
- b. What is the throughput of people in EFI - what is the gain rate, what is the burn rate, and what is the stock
- c. Who comes in and excited about the concepts, but then get lost
- d. If Task View people are interested in visualization - then work on visualizing as a concrete task. Would be good for a group that is jazzed about
 - i. Where is the energy and where can we dive in?
- e. GLEON working groups are working toward a project. Folks get involved coming out of that
 - i. What is the next investment

- ii. What are the things people want to work on that fits with their work already because people don't have time to work on things not related to their interests
 - iii. Do we need static working groups? Perhaps they sunset. What ends the working group? Working group ends once they finish a task?
 - iv. When starting up a new project - put out a call about new opportunities to participate. Could have onboarding
5. Forecast visualization challenge pitch (this came up on the Sept call, think about if this is something the group wants to move forward with)
- a. Leaving this as a placeholder, but not quite ready to move forward this, right now.
 - b. Update from Jody on November call - we are working on details for an RCN in-person meeting next June and will be putting out a save the date in the next month. The plan is to have people/groups propose topics for concrete activities to work on during the meeting and then have people sign up for those activities prior to arriving at the meeting (people can also change groups/activities once they arrive). But this idea, would be a great thing to propose for the meeting.