

August 11, 2022 Joint Methods & CI Working Group Call

Attendees: Matt Brousil, Quinn Thomas, Jody Peters

Agenda/Notes:

1. Poll to schedule calls in September to December
 - a. Make sure your time zone is selected

2. Visualization Task View - so close...
 - a. Link to [the New Copy of the Doc](#)
 - b. Uncertainty Quantification & Propagation, Modeling & Stats and Workflow Task Views to use as Reference: <https://projects.ecoforecast.org/taskviews/>
 - c. [Task Views GitHub repository](#)
 - d. Libby pushed updates to fix the 2d histogram and installed the arrows package to download the NOAA data to fix the correlation plot. We also think she may have figured out the stringi package issue with the GitHub actions. There is a renv.lock file that references the packages used and that has not been updated yet. Jody will work on this early next week.
 - e. Jody is checking in about the HOP example Libby was going to work on
 - f. Check in with Mike: Mike and Carl had a conversation about cloud-native file format options in Slack. Mike - is there anything from that conversation that you think should be followed up for the Task View?
 - g. Quinn uses Quarto for NEON Challenge and it is pretty easy of
 - i. Rename from Rmd to Qmd and change chapters.
 - ii. The GitHub Actions is easier to use. Could replace the current one with what Quinn has

3. Forecasting Standards Update
 - a. Most recent version 0.4 of the Standards
 - b. Standards has stabilized and has different file formats from previous versions.
 - c. Don't have a lot of people submitting this round so it has been a good time to do a big update
 - d. The terrestrial, phenology, and aquatic themes are near real time.
 - e. To add a new site, just need to change one line on a csv that sits on GitHub
 - f. Beetles is real time in that you are forecast the future - so no one needs to watch it
 - g. For Ticks, John was working on having one month of feedback, but it is not a true forecast. So the other 4 themes don't have the concept of a round. They are always available to submit a forecast to. So when we roll over to a new year it doesn't change anything.
 - h. But for Ticks it is the only one that requires changing and making updates for new rounds

- i. For Matt - it has been a big thing to wrap their head around the time lag and considering how to incorporate the weather data. If it could be streamlined to match the other themes it would be useful for Matt's group.
 - i. His group is close to making a forecast
 - ii. Right now it is predicting a month into the future.
 - iii. Would it be better to predict a year into the future. The beetles are forecasting weekly a year into the future.
 - iv. If doing short term forecast, working with more recent historical data, whereas iwth the year out using more. Matt has more interest in the shorter term forecast. Think it is more tangible and more real world application
 - v. Time step vs horizon
 - vi. Longer term forecast would be predict in 2 week intervals.
 - vii. Like the idea of weekly or 2 week intervals
 - viii. Do forecasts get re-evaluated? Yes, could submit a forecast any day. In theory could imagine stopping all forecasts next Aug 11. So now you submit a 365 day ahead forecast or 52 week forecast. Tomorrow, it could be 364 day and the next is 363 and as you get closer you are revising. This is sort of like a 583 forecast. This is what I thought a forecast would be 1 year out, 9 months out, 2 months out, 1 day out.
 - ix. Phenology gropu has looked at timing events (e.g., when greenup occurred) and looked at how good the forecasts were 1 day before or 1 week before the greenup. It works for phenology becuae you know when the event happened - when the greenup happened
 - x. Ticks are more tricky becuae you need to forecast now before going into the future. Becuase now is technically a forecast becuae the ticks have been collected, but we don't have the data for a year out.
 - xi. What is a model for the Challenge?
 - 1. If using fable package - fit mean model and random forest model. Would need to take df and filter it by each model and write a csv for each model
 - 2. If evaluating models for sites differently. Fitting a model to all the sites and seeing what worked better for each site.
 - 3. If model X works better at site Y.
 - 4. Then can have model J for sites A,B,C
 - 5. Dot model in fable is a team
 - 6. Better off submitting all the sites for a model and submit both and the analysis after the affect shows what model works best
 - 7. Forecast models are numerical guidance to a forecaster. If a forecaster wants the best model, may want to blend models. So there could be an expert opinion model that is a blend.
 - 8. Whole point is to work through sticky issues. Submit your best guess and if it is blended models that is fine as long as you can describe it.

9. Think about automating. If the GitHub Actions is the limiting factor, we can help with that
 - xii. Talking about doing a hackathon or have help hours or getting together for a day to help with some of these issues. Feel there are a lot of 80% forecasts and how to get the last 20%. If you have a cool model that is ready to go and then figure out how to get it automated.
 - xiii. Have lots of grad students interested but need to learn so this could be helpful.
 - j. Future topics for this group to consider in regards to Standards
 - i. What tools can we develop to help teams create and generate the metadata
 - ii.
4. Workshop Proposal
- a. Identify gaps in CI/methods for forecasting and put forth NSF proposal for a workshop/series of workshops for \$50-100K. Want something modular and interoperable. Want to pull in people with CI and computing resources
5. Data Ingest, Cleaning, and Management Task View
- a. From June call - Brittany can contribute to remote sensing, Matt works on this topic all the time, and Kelly was interested in this
 - b. Matt might be interested in leading/co-leading. Can't do it until mid-August to early-September.
 - c. Jody happy to help with co-leading. Brittany will look through
 - d. Don't follow up on August call, but reassess how Matt is doing with his other projects and then before September try to go through the outline to clean it up and then run it by the group on the September call
 - e. Brittany gave a presentation to R ladies on parallel processing and big data so could add to that
6. Forecast Workflow Example
- a. [Example from NEON Challenge](#) - uses Aquatic Ecosystems
 - b. Mike's carbon cycle forecast model he prepped for Flux Course - uses NEON flux data from Niwot Ridge
 - i. <https://github.com/mdietze/FluxCourseForecast>
 - c. What would be nice to update or add to make the on ramp to creating forecasts easier?
 - d. Matt has been working on a workflow for the Tick Challenge. Has a workflow for that is built with targets. Has it documented on OSF right now.
 - i. <https://osf.io/e4uhy/wiki/home/>
 - e. Something to consider moving forward for the workflow examples
 - i. What is the best to catalog the example workflows? You can see examples in GitHub, but they are often hard to follow

7. [NEON Ecological Forecast Challenge](#) CI Update