June 9, 2022 Joint Methods & CI Working Group Call

Attendees: Matt Brousil, Brittany Barker, Mike Dietze, Jody Peters, Kelly Heilman, Ben Toh

Agenda/Notes:

1. Debrief from Virtual Conference

2. Visualization/Decision Support Tools, User Interface Task View
    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. Jody has pushed the updated visualization.Rmd to the Task Views GitHub repository
    d. Status of the Task View and remaining needs
        i. What was done:
            1. Confirmed that we can use the PNAS figure for educational use (and we have Joel Smith's okay as well)
                a. From PNAS: Permission is not required to use original figures or tables for noncommercial and educational use (i.e., in a review article, in a book that is not for sale) if the article published under the exclusive PNAS License to Publish.
            2. Tables 1-3 are now at the end
            3. Text for the High-level Reactive Tools is updated
                a. Ben is still working on details for Observable
            4. Libby's code was added - need to check
                a. Fig 3.2 - the hexagonal heatmap showing greenness and redness for 6 NEON sites
                b. Fig 3.11 - the combined scatterplot matrix and correlation matrix for 5 weather variables at Oak Ridge. Jody has a inquiry out to Carl Boettiger to see if he has suggestions for the error when trying to get the driver data
        ii. What Still needs to be done
            1. Have a few citations that are missing, mainly from the Uncertainty section
            2. Jody is getting error when building the taskviews bookdown page about failure to install stringi package: https://github.com/eco4cast/taskviews/actions/runs/2469613650. She will work on this
            3. When the taskviews bookdown page is updated, need to check the links to internal sections
            4. Work with Libby and Carl on Fig 3.2 and Fig 3.11

5. Do we want to include text setting up or labeling the example plots in the Static section? Do we want to show the code for the R setup for Libby's figures?
    a. Provide a link this code so people can get it if they want to run it themselves. But it is too much code for the current task view. However, if we can hide the chunks of code, then maybe we can include it here. Matt has some code to hide chunks of code for markdown that will probably work.
6. Add text about Observable in the High level reactive section
7. Reference Melissa Kenney's talk from the EFI Virtual Conference on "Improving Decision-maker Usability of Forecast Data Products" someplace in the Task View
8. Next steps
    a. Jody will work on cleaning up the Task View and work on these few remaining needs
    b. Then send it Matt and Brittany to take another look. Then will send it out to the whole group for review before sending it out on Slack, Newsletter, Twitter, etc

3. Discuss ideas for the Project(s)/Activities for the group to work on next and who will lead
    a. Ideas (see point C below for a fuller description of each idea)
        i. 1. Data Ingest, Cleaning, and Management Task View
        ii. 2. Identify gaps in CI/methods then work on small doable chunks to fill those gaps
        iii. 3. Projects that make the NEON Forecasting Challenge better (this can be for synthesizing or visualizing forecasts, making onboarding for new teams easier, etc)
    b. Notes from the June 9 discussion:
        i. For Numbers 2 (gaps) and 3 (Challenge) below - need to think about breaking things down into small management chunks as a working group. Want to avoid burnout. The Task Views were manageable chunks to push out something useful for the community. It isn't another chapter for someone to work on.
        ii. Brainstorm ideas for what would fill gaps that would be scalable. Then think about how to get it done (do it over a series of WG calls, tackle a number of things at once with a hackathon/a meet up at ESA, other things to grow the community etc)
        iii. Also, see notes below, particularly in points c(i) and c(iii)

    c. 3 Ideas Previously Proposed
        i. Data Ingest, Cleaning, and Management Task View

1. Brittany can contribute to remote sensing
2. The Task View outline is currently set up to go over general principles and tools and common types of data that ecologists often work with
3. This is a topic Matt connects with and spends much of his time working on. The data cleaning and management is the thing that most people spend tons of time working on and the more clean and organized the data, the easier the analyses are.
4. The Task View outline has areas where we can provide nitty gritty details about how to clean the data and ingest. But thinking about the gaps and needs for the community - want to share with the community how to start with data ingest and then move to build an operational forecast.
    a. Are these two separate ideas/projects or would it be useful to work on these in parallel?
        i. The group could work on these two things in parallel
    b. We do need people willing to take the lead - don't need people to take the lead on the whole document, but if there are chunks of things people are interested in, then that would be great
5. Would we want to have an annotated forecasting workflow with breakout with more specifics in some area? What options would we want?
    a. Data ingest and cleaning - can have some examples of how to deal with big data. Right now the outline looks like a list of best practices and list of tools. The discussion about gaps seemed like something to have a document that identifies gaps and would also provide fodder for a mini workshop with examples in it.
    b. Think this group could easily work on identifying small gaps. Then scaling up - would like to think about a broader goal for tools and interoperability. There is interest for this broader goal. EFI has had a number of interagency calls with individuals from across multiple US agencies. CI and stakeholder engagement topics are the big issues that keep coming up in those calls. Think we are convincing the people who are funding forecasts (NASA, USGS, NOAA) that having a shared infrastructure is needed so each new proposal funded doesn't require the forecasting workflow to constantly need to be reinvented.

i. Would love to have something that provides examples of modules and how they would be interoperable
ii. If we can make a pitch for a workshop to bring people together who have the cyberinfrastructure and computing resources to move this forward. Think NSF would be on board for this.
iii. $50K vs $100K. $50K doesn't need to be reviewed by a panel, can be approved by the PO. Up to $100K also do not need to reviewed by a panel (but needs to be reviewed by other POs)
iv. Do workshop funds provide for salary for someone who is not tenured? Think we can do that.

c. Next steps for the next call -
i. Outline workshop proposal
ii. Example workflow - Mike will be teaching about forecasting (focus on the C cycle) for 1 day at Flux course (the goal is to get students up and running for creating forecast in one day of instruction)
iii. Here is an example workflow from the NEON Challenge - using the aquatics challenge. This could be a good example to start with - what needs to be updated, where could it be made more general?

ii. Identifying gaps in CI/Methods and brainstorming how to fill the gaps
1. Notes from Mike from fall of 2021
a. could have some useful discussions about identifying gaps and bottlenecks in current methods and tools and ways we might think about moving forward on identifying and addressing those (e.g. surveys, workshops, hackathons, group projects, virtual panels / webinars, etc). This is also motivated by some recent interagency (NOAA, NASA, CDC) roundtables that EFI has been sponsoring where standards and community CI are emerging as priority items for the federal agencies (caveat, that doesn't mean there's any funds). It is also relevant to the next round of the EFI NEON challenge and the upcoming EFI2022 "all-hands" conference

iii. Visualization synthesizing the NEON Forecast Challenge - idea from Carl and Quinn
    1. Additional suggestion/discussion from June 9 call - if there are any places the community wants to dive in and enhance the Challenge CI (not just visualization) that is very welcome
        a. We have switched to cloud native architecture - now that we have made that switch there are things that can be enhanced
        b. Help teams get on board with the forecast challenge. Have null forecasts for each of the themes, but not sure if there is a good up to date worked example/primer/repo that people could fork to start from for the generic forecast. - what are the chunks of code that are useful to help people get up and running on creating forecasts - this fits nicely with defining the gaps listed above
        c. Example - looking up the last forecast and figuring out where to start the next forecast.
        d. What are the universal steps that are needed for any type of model?
        e. Show the different ways to do things - part of the education
        f. Earlier Task View on workflow provide resources for workflows for forecasts, cron jobs, action.
            i. Gap between documenting it in the Task View and reducing the slope of the on ramp for people to be able to create forecasts
            ii. How to make the on ramp shallower and how to reduce redundancies?
            iii. If we have a more clear understanding of the steps involved adn have basic interoperability then can allow people to reuse what others have done.
            iv. We are limited in the field because there is non trivial computer training needed and for keeping forecasting systems running. So if we can share that burden across the community that would be useful
            v. Reduce maintenance cost of running the systems, reduce slope of onramp into the systems, and reducing redundancy so people aren't continually reinventing the wheel.

vi. Operate in the mid-size (small - run forecast of this population once a year - can run it once a year with 40 lines of code vs huge - groups with dedicated supercomputers adn CI staff to predict one thing. EFI is in between. There is no center to operationalize them.)

2. Visualization notes: Quinn and I were talking about how we'd love to start engaging this WG more directly with methods/CI in the neon4cast challenge. In particular, now that we have a decent pipeline for gathering in forecasts and publishing scores, there's a lot to be explored with visualization. The current flexdashboard is maybe an example of this (https://projects.ecoforecast.org/neon4cast-dashboard/phenology), it shows a range of possible visualizations and summaries but is kinda all-over-the-place, probably few people visiting the site can make much sense of it. For an audience that knows ggplot and wants to learn/practice more advanced tools like interactive visualizations and dashboard layouts, this could be a good entry point. It would also be fun to get more folks engaged with the data coming out of the challenge, and using visualization to explore interesting questions about forecast horizons performance according to different metrics, visualizing uncertainty, etc etc.
3. We could get people started in creating simple interactive dashboards with minimal R code showing an overall team rankings or a single forecast of a single team. I imagine then that WG members might work individually or in groups to generate some visualizations, and/or critique and tweak existing visuals (and supporting text!) so we slowly move towards dashboards that really tell a story about these forecasts.
4. Anyway, just wanted to bounce the idea off you. I don't want to derail the ongoing work of the group, and maybe this idea is terribly impractical, as you know I'm prone to being carried away with impractical ideas!

4. NEON Ecological Forecast Challenge CI Update