September 16, 2022 Joint Methods & CI Working Group Call

Attendees: Mike Dietze, Jake Zwart, Jody Peters, Brittany Barker, Leah Johnson, Ben Toh
Regrets: Matt Brousil, Hassan Moustahfid

Agenda/Notes:

1. Announcement - the Statistical Methods Webinar Series co-hosted by EFI and the ESA Stats Setion will begin on Oct 3
   a. October 3, 2022 – Spatial Occupancy Models and the spOccupancy R package; Andrew Finley and Jeffrey Doser, Michigan State University
   b. November 7, 2022 – Analysis of Bioacoustic Data; Marcelo Araya Salas, Universidad de Costa Rica
   c. December 5, 2022 – Structural Equation Models and the piecewiseSEM R package; Jon Lefcheck, Smithsonian Environmental Research Center
   d. January 9, 2023 – Zero Inflated GLM and GLMM, Alain Zuur and Elena Ieno, Highland Statistics Ltd.
   e. February 6, 2023 – Spatial Modeling in Ecology; Marie-Josée Fortin, University of Toronto
   f. March 6, 2023 – State-space Models and the Template Model Builder (TMB) R package; Marie Auger-Méthé, University of British Columbia
   g. April 3, 2023 – Bayesian Stable Isotope Mixing Models and the MixSIAR R package; Brian Stock, NOAA
   h. May 1, 2023 – Integrated Species Distribution Models; Neil Gilbert, Michigan State University

2. Visualization Task View - so close…
   a. FINALLY! The text/figures are up on the Task View Bookdown site! https://projects.ecoforecast.org/taskviews/visualization.html#Uncertainty-Visualization
      i. Huge shout out to Libby who figured out we needed to update the renv.lock file which points to the packages used!
      ii. Next step - have the group review it.
      iii. Jody needs to go through and add links to different sections
      iv. Would like to send this out in the Newsletter that will go out on Sept 27
      v. Jody is looking into moving from bookdown to quarto
         1. Similar to bookdown and pretty similar file structure
   b. Old resources from the Task View
      i. Uncertainty Quantification & Propagation, Modeling & Stats and Workflow Task Views to use as Reference: https://projects.ecoforecast.org/taskviews/
      ii. Task Views GitHub repository
      iii. Libby added the HOP example

    c. Visualization Tools Hassan found recently at a NOAA EDMW meeting
        i. [4DVD](#) very interesting tool. I checked out this tool and it looks really interesting.
        ii. [GBIF occurrence density map.](#)

3. [Data Ingest, Cleaning, and Management](#) Task View
    a. Matt is not able to make today's call, so postpone this until October's call

4. Forecasting Standards and [NEON Forecasting Challenge](#) CI Update
    a. Most recent version 0.4 of the Standards
    b. 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
    c. Challenge CI is evolving and getting more stable - have system that tprocesses targets and inputs for every NEON site for all themes (except for ticks which is done only for NEON sites where the target tick species are present)
    d. Updates to helper function to make those more usable for teams
    e. Substantial update on example of the [workflow from the NEON Challenge](#)
        i. This example sets it up a GitHub action so that the forecasts are set up with a cron job so the forecasts don't have to be run on personal computers and are automated to run
        ii. Building docker container which can be used to create forecast that has all the forecasting packages built in. If anything custom is needed, you can still add that
    f. Standards:
        i. Could provide summary stats of ensembles, but couldn't say what type of distribution it follows
            1. Standards team has been talking with the fable team - now the new standards include the distribution names as they are used in fable
        ii. Changed some variable names to be consistent with [the STAC](#) community
        iii. Mike is working on version 0.5 with these updates
        iv. Are the standards interoperable with climate forecasting standards?
            1. If we change from time to datetime it will no longer be what cf calls it
            2. Need to check how the STAC changes affect the cf convention - Mike to reach out to Quinn/Carl
        v. Looking for champions - this will replace the design team role
            1. We are still working on the detail for the champions, but wanted to reach out to the group to share the opportunity
            2. Manuscripts can be led by anyone, but champions would have the opportunity to participate or lead in manuscripts

3. Forecasting blitzes - specific times where we encourage many teams to submit forecasts for specific periods of activity (e.g., spring greenup, tick emergence, etc)
4. Threshold for swag as a reward for participation - if you submit 25 forecasts you get $25 in the EFI swag store
5. Currently the challenges are focused on NEON data for the 5 themes. There is interest in expanding the challenge, but we don't have funding to set that up yet. LTAR, TERN, LTER are all communities we have reached out to. But would take effort to get the data from those organizations into the Challenge CI
6. Idea from Jake - think there could be additional aquatics sites that could be spun up like USGS gauge sites.
    a. Good to follow up with Quinn - if people want to add sites to the challenge and have interoperable data, is there an approval process, or is it just building the data so the targets get built?
    b. How do I apply this to my own site where it has frequently updated target site.
    c. Good for someone in the RCN to write up documentation to apply this to new challenges - how to stand up the stack to get it running. Mike got it 90% working when he tested it out one day, but had to ping Carl a couple of times.
7. Coming back to the Task View concept - the Challenge CI is all cloud based. Not sure if any of our existing Task Views provides an introduction to cloud native data. Should this go into Data Task View? Or bring into one of the other Task Views?
    a. Think this will fit well with the upcoming Data Task View - Mike will add a bullet point to the outline for this task view
    b. Could bolster the text in the containerization subsection in the first Task View where we talked about the concept but didn't say this is great from going from your laptop to the cloud
    c. For other Task View components, not much is focused on using the cloud - it is more general. But could mention it in the modeling maybe talk about resources needs and different options
    d. The way the cloud native formats data under the hood is advantageous to forecasting. You can add new tiles for the last week of data without having to update the old data. The system perceives it as the data is bigger, but it is just another file that gets added to the system
        i. This is better than physically appending rows on a file

8. The image Quinn was working on is published - https://github.com/eco4cast/neon4cast-rocker
   a. The actual container gets built. His example forecasting workflow uses this container. The whole workflow runs in ~2 minutes

5. Forecast visualization challenge pitch
   a. using new or archived EFI forecasts
   b. scored based on the correct understanding of the forecast predictions and characterization of uncertainty and risk assessment
   c. Many ways to visualize forecasts and uncertainty. Challenge provides an opportunity with all the forecasts that are getting submitted
   d. Melissa Kenney would be good to reach out to about how to score this type of Challenge
   e. Social Science group has had this conversation. Don't have an active timeline for how to do this.
   f. Sticking point - how to automate a scoring system for this. How to judge how efficient visualizations are
      i. Do more than ask if people like visualizations. Could make a quiz that asks people to evaluate visualization and see what they grasp from the
      ii. Brittany works with someone who works with end users to get feedback from users.
         1. Depends on what are the end goals - are we trying to improve visualizations so they can make decisions?
         2. Or is it a research study to make it the best visualization? If it is this, then need to have a big enough sample size of people to evaluate the visualizations.
   g. Idea from Jake about what the goal could be is to come to the right decision. Not sure what the decision would be. Could pick out certain time periods where people would need to make a decision or have a scenario that people need to make a decision to make and ask people what decision they would make with the decision
      i. This would be more educational rather than making management decisions
      ii. What are the NEON Challenge forecasts being used for?
         1. If to manage, then want to ask the managers what info they need to manage the ecosystem
      iii. Brittany is using a web app to forecasts infection risk for boxwood blight
         1. End users would be the people going out (Oregon Dept of Ag) looking for outbreaks or nursery managers looking for infections.
      iv. Need identification of early adopters to say - if I had forecast like this for my site, how would I use this to make decisions
      v. Most NEON sites are not intensively managed and NEON has no control over the way the sites are managed

      vi.   So then it would be on the specific type of research questions people want to answer
- 1. Then the challenge will be to get enough people to score those visualizations

      vii. Step 1 - identify the questions we want people to answer with the visualizations. This will require discussions within the different theme areas

      viii. Carl and Quin have been working on this dashboard to compare across forecasts and visualize scores: https://projects.ecoforecast.org/neon4cast-dashboard/phenology
- 1. Carl has been wanting to get input on the dashboard and help thinking through what questions the dashboard could be used

      ix. Could overlap with this idea - could run a competition to get visualization ideas that could then be incorporated into the dashboard

      x.   Expect this will need a champion - will need someone to take the lead on a WG to outline the steps and recruit people into working on those steps.

      xi. There is a Twitter visualization challenge that happens that has different themes for every day in the month of April. Could include uncertainty in there.
- 1. Could incorporate into this Challenge - we have this dataset that is free for anyone to participate
- 2. Or could use social media to get sufficient feedback on questions about visualization
- 3. Could reach out to people working in similar systems - federal agencies, USGS, US Forest

6. 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
      i. NSF doesn't need to send out these types of proposals for review
      ii. Mike has pitched a roadmap white paper that the community agrees to what community CI should look like - how it is designed, what are the pieces it should include.
      iii. Then could ask across agencies to pay to support that CI. No one agency wants to pay the bill for this. NSF can't operationalize
      iv. Jake has been talking with the interagency working group. This is an issue - there aren't forecasting centers or offices not tied to the National Weather Service. So it is hard to have sustained funding for operational products
      v. At USGS there are more projects that want to do short term forecasting. Over the next decade there will be more need for this, but don't have a plan to scale up

vi. People at NOAA already implement but it is boutique since they don't have a forecast center

vii. There is a WG within ICAMS that works on CI - this is the big issue in that community as well

viii. US NPN used to get funding. Don't anymore because of changeover in leadership.

ix. Key word = boutique. If you build community tools, it will cost less since everyone will not need to recreate forecasting workflows

x. We need a plan for the CI - that is the pitch for the workshop.
1. Get $50-100K from NSF and can work to get NASA/USGS to put in that much as well. Want to get agency buy in so that they will pay attention to the

xi. Need a champion - don't think it would be hard. Think it would take a couple of weeks to pull together
1. Think about the timing for this adopted
   a. This upcoming year could be good from USGS
   b. There will be a meeting with USGS and NOAA around modeling efforts that include forecasting
   c. There will be a forecasting strategic plan written in 2023 for USGS
   d. NSF and NOAA are having their own pairwise conversations

xii. Next steps - open a shared doc on another call to collaboratively work on an outline of the steps and the pitch for the proposal

xiii. Then have a champion(s) shop it around to agencies

7. 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 rae often hard to follow

8. [NEON Ecological Forecast Challenge](#) CI Update (see update in point 4 above)