

June 14, 2023 Joint Methods & CI Working Group Call

Attendees: Mike Dietze, Quinn Thomas, Jody Peters

Regrets: Jessica Burnett

Agenda/Notes:

1. Mike is working on updating his material for Flux Course to teach ecological forecasting using the NEON flux Challenge
 - a. Setting up GitHub actions and setting up the yaml for metadata so that if people fork his repo that they won't submit forecasts for his model.
 - b. Hasn't used MODIS as a constraint, planning to use it as part of the class project, but if he can get it up and running himself then will include it.
 - c. Goal is to get his model officially submitting and expand out past Niwot Ridge.
 - d. Thinking about restart issues. Stache to GitHub in a file that overwrite's itself - if you use variables that don't use keywords that are variables for the Challenge, then not sure if it will be kept
 - e. Bucket version requires user to have access to a bucket. So would staching on GitHub work for a simple model? Once a day a file doesn't get appended. But Mike is appending and it is a binary model. Could create a folder that a new file gets added to. These would be tiny files so would be okay
 - f. For FLARE - do a 5 day jump back even though there is a 2 day latency, but want to give enough of a buffer
 - g. Mike does have it set up to query stage 2 and stage 3 met forecasts which is cool
 - h. Mike has whole day to teach forecasting and is starting with the assumption that people are starting with knowledge of R.
2. Quinn's update - working on incorporating TERN NEE and LE forecasts which gives a test run for "forecast challenge in a box" has jobs for scoring, downloading forecasts, building dashboard, etc and has it set up to run in GitHub actions (tern4cast).
 - a. Combined documentation and dashboard into a single page.
 - b. In the S3 bucket, it is not merged into the NEON forecasts. All the forecasts are the same, but they are kept separate. A person would have to submit to both NEON and TERN.
 - c. To be able to replicate this, the person setting it up would need 2 S3 buckets and be aware that
 - d. It will eventually be set up to update to the STAC catalog for the Challenge
3. CI Workshop Proposal Update - Cleaned up version to collectively develop ideas
 - a. Goal: Identifying and filling gaps in CI/Methods for forecasting. Bringing together people from gov't agencies, academia, and private sector/NGO
 - b. Planning to have 2 meetings, one in November 2023 that focuses on the cyberinfrastructure and another 6-8 months later that focuses on research to operations.

- c. NSF OISE no longer has funds for workshops. But AccelNet was suggested. Here are the details about that.
 - i. You might take a look at the [AccelNet solicitation](#), which focuses on networks of networks. AccelNet is designed not to support research per se but to support the activities that allow researchers to come together and collaborate more effectively across networks through things like roadmapping exercises, workshops, student exchanges, etc. The Design track is for teams that are in the formative stages of their work. The AccelNet solicitation is under revision but the bones of the program will remain largely the same. You'll want to look for the revision once it's posted in 2 months or so, but for now the existing solicitation is a good reference.
 - d. Still waiting to hear back from NSF DBI
 - e. NOAA funds are secured - thank you Hassan!
 - f. Working with Jake Kritzer from NERACOOS to work out details for hosting the first workshop at NERACOOS and to submit a proposal for additional NASA funds

- 4. Forecasting Challenge CI Updates
 - a. STAC [prototype](#) for making the Challenge forecasts searchable
 - i. Austin is working on adding keywords, description, and NEON sites to the STAC descriptions
 - ii. Can find the embedded metadata on schema.org which allows the forecasts to be searchable at a single command line.
 - 1. Will allow people to search for a person, search for "linear model", or a certain target data (e.g., "dissolved oxygen")
 - iii. Next steps - how to incorporate nuanced model descriptions including how you represent uncertainty
 - b. Issue that came up in the Theory working group call
 - i. The Forecast Challenge [gives people 2 ways](#) to provide metadata. 1) Registration is required where people fill in info on a form that asks about the model and the uncertainty. 2) People can optionally submit an xml file. But the generate_metadata function in the neon4cast package does not work.
 - 1. Quinn will update the instructions since the generate_metadata is out of date
 - ii. If Theory group wants to create xml from scratch, then the group can.
 - iii. The helper that goes from yaml to xml is out of date.
 - iv. So continue to push people to make updates via the registration form

- 5. Unconference preparations and sharing ideas posted in for the Unconference <https://github.com/eco4cast/unconf-2023>

- a. There were 24 project proposed, we think the group will only have time to work on about 10 projects, so there may be opportunities for this group to follow up on project that get developed during the Unconference or that

6. Leaving this in as a reference for something we may want to pick up on after a reboot from the Unconference - Brainstorm educational activities that could be developed/promoted by this group
 - a. Longer term ideas for educational opportunities related to CI/methods for the broader EFI community
 - b. Stats seminar-like presentations or how to use the Stats webinars to bring people into EFI activities
 - c. Reading group?
 - d. Think about this after the Unconference and with plans for a reboot

7. Other Resources/Previous Project Ideas
 - a. Data Ingest, Cleaning, and Management Task View
 - b. Discuss papers that cover existing cyberinfrastructure or stats methods
 - c. Organize, present, or view presentations about topics in cyberinfrastructure or a stats method