April 12, 2023 Joint Methods & CI Working Group Call

Attendees: Jon Borrelli, Jake Zwart, Jody Peters, Mike Dietze, Quinn Thomas, Jessica Burnett

Agenda/Notes:

1. Updates and Next Steps
    a. CI Workshop Proposal Update - Cleaned up version to collectively develop ideas
        i. Goal: Identifying and filling gaps in CI/Methods for forecasting. Bringing together people from gov't agencies, academia, and private sector/NGO
        ii. Currently shopping it around to see if there is buy in and talking about timeline to see what a realistic time to host this - sounds like Oct/Nov 2023
        iii. Need input on content and outcomes
        iv. Would like input/feedback on proposing this to agencies
        v. For NSF - will send this unofficially to get feedback and NSF POs will tell you what they need you to send officially
        vi. Think $150K is high for 30 participants
            1. Jessica will work on budget with the NCTC pricing
            2. This amount would be good for a series of meetings
            3. If pitch it as an initial meeting and send people home with homework and then have a second meeting to congeal around and get something concrete
                a. What is the turn over? Would it be a feature or a bug?
                b. We will use the outcomes of the first to recruit people to the second.
                c. Mike - we may have text like this from the Synthesis Center proposal
                    i. Turnover for people who decide not to come back and targeting additional people
                    ii. You can sell the turnover as a feature. We will target additional invites from what we learn from the first meeting
            4. Size - 30 is good if you have breakouts. Too big if you expect everyone to focus on a specific task
            5. Would like feedback on the participant list. The diversity is lacking.
                a. For each person if you brainstorm, this is the perspective they will bring and if there is too much overlap
                    i. Have categories of expertise, background, and fill people in
                    ii. From the list of potential participants - NASA has biodiversity overrepresented, ecosystems and CI underrepresented. Is anyone from DACC available? Bruce Wilson

1. Breelyn from USGS
2. Jessica will think about this more

iii. Peter McCartney is retired from NSF (or very soon) - will be good to ping to see who should attend instead of him
1. Mike talked to him about this as recently as the Sept LTER meeting

vii. Turn around for commitments from agencies - when to get initial commitment?
1. Most agencies like 6 month lead time
2. Will small amount like this could get away with
3. Shoot for beginning of early FY
4. If we run into getting funds that are asynchronously available, then getting the request in early will allow for the most time to use the funds over the year
5. **Put out the informal request as soon as we can to get feedback on what is actually needed for an official submission**

viii. Jake/Jessica - run it by USGS, NASA. Chris Brown can help with NOAA and Hassan Moustahfid is looking at it as well. Mike will check in with NSF

b. The Forecasting Standards Paper was accepted to Ecosphere

c. Forecasting Challenge CI Updates
i. Carl has been working on being able to download the climate forecasting system from Amazon as a tool that is part of the R packages on ecoforecast GitHub repo
1. A bit wacky - the horizon isn't consistent from the 6 hourly (goes out 5-7 months) but calls it a 9 month product. Not all ensembles go out the full length

ii. Working on getting v11 of the NOAA GEFS which is also in Amazon. This will let people go back to 2015. Intention is to not download all forecasts, but the first 6 hours - first time step of the NOAA weather forecast in order to calibrate a model. Then look to see if it is possible to go back further in time.
1. Can add new sites easily to get the weather forecasts for new sites/non-NEON sites

iii. Working on STAC metadata catalog for the NEON forecasts. To make them discoverable

iv. Lots of aquatics forecasts coming in

v. Not much for phenology - wonder where phenology people come together

1. Think about reinvigorating this group. Think about training people who previously submitted forecasts to set up their forecasts to be auomated
2. Could tweak a pitch to have nature climate solution language

vi. Need champions for the non-aquatics themes

vii. For terrestrial - wonder if people are used to calibrating their process models and the idea of calibrating to 47 sites is daunting/hold up

1. Think the hang up is running the 47 sites iteratively to set up the workflow to submit every day/month and why to do that.
2. Would help for terrestrial side ot have a worked example with a non-trivial model that people can emulate
3. Even people if can run in GitHub actions (which some won't be) it will be the issue of getting it into the container to run the forecast
4. Mike's example from his book can easily be set up in GitHub actions - took particle filtration approach and calibrated it to eddy covariance. It is 90% ready, just need to get the GitHub actions set up. Think he can get it ready for this year's Flux Course

d. Translation Database - the Translation working group is thinking about match making in EFI to make connections across the group. Chris Brown is taking the lead on pulling together information with input from the Translation group. Chris reached out to Jake and Jake provided a number of ideas (see the list below). Does anyone else in the group have experience with any other matchmaking databases?

i. You could think of it as a research based thing instead of person-based. Then maybe ICPSR is a good example https://www.icpsr.umich.edu/web/pages/ICPSR/index.html e.g. what are you studying, and then find the people studying that

ii. civicrm (https://civicrm.org/) is a tool that manages people and membership kind of information and is generally a constituent relationship management system. Something like this or using keywords in a Google search for a solution could be useful

iii. here's a useful example of what I think you're trying to do from natural hazards SSEER - CONVERGE | Natural Hazards Center (colorado.edu)

iv. There might be some inspiration from human dimensions https://doi.sciencebase.gov/hd/ group too but no specific examples of that yet

v. We had previously started a similar initiative for matching scientists with appropriate skill sets within USGS Water Science Centers, as part of personnel budgeting reports. The system allowed individuals to create a profile by adding their active directory name and email to a table and selecting from a list of skills that had been organized into higher level groups, sub-groups, and lists. However, there wasn't much buy-in and the initiative did not progress far due to changes in gov't policies and technology. Maybe not too helpful but nice to know others were thinking about this

vi. This might also be useful https://github.com/vivo-project/VIVO

vii. From Chris: Potential to connect with NOAA's National Water Center

viii. NEON has adopted software for managing its communication with the community (e.g., one person would do outreach to one group and another would connect with another group, but there were issues in communication breaking down when there was staff turnover)
    1. Jody will reach out to Christine

e. Does anyone have experience with protocols for registering projects to let the community know what is being worked on, how to connect, and avoiding duplicate papers?
    i. The RCN Steering Committee talked about this and it is something for the EFI Steering Committee to discuss, but if others have experience with doing this for other groups, it would be good to compile what has already been done
    ii. GLEON
    iii. Ameriflux

f. Stats Seminar Series - ideas for R packages related to forecasting that the group would be interested in hearing about?
    i. Nick Clark offered to present on mvgam - Dynamic GAMS
    ii. Want to find presenters that are early career (grad/postdoc) and from underrepresented backgrounds
    iii. fable package - Tidy forecasting in R. Check to see if Australian group knows Rob Hyndman - Applications of Forecasting
    iv. Tidymodel framework
        1. How to guide: https://github.com/frec-3044/machine-learning-template
    v. From Mike - Stuff in the Space Time book reading group did. Think about Chris Wikle's former student, Toryn Schafer as a potential presenter
    vi. https://greta-stats.org - GPU based training of MCMC models trying to overcome scalability of large datasets
        1. Sort of like NIMBLE where it seem powerful but not enough people use it to have lots of documentation and a whole community
        2. Have lots of examples - but would be good to see a forecasting model

g. Brainstorm educational activities that could be developed/promoted by this group
    i. Short term - any additional Unconference ideas? https://github.com/eco4cast/unconf-2023
    ii. Longer term ideas for educational opportunities related to CI/methods for the broader EFI community
    iii. Stats seminar-like presentations or how to use the Stats webinars to bring people into EFI activities

iv. Reading group?

2. 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