January 18, 2023 Joint Methods & CI Working Group Call

Attendees: Quinn Thomas, Carl Boettiger, Mike Dietze, Leah Johnson, Jody Peters, Jessica Burnett, Chris Jones, Jake Zwart

Agenda/Notes:

1. Announcements
   a. [Applications] are open June 21-23, 2023 [EFI Unconference] at NEON HQ in Boulder, CO Save the Date. **Applications are due February 1**
      i. Goal: Bringing People Together to Do Forecasting: Training, Technology, Theory, and Translation
         1. Participants will work together to produce products they want to work on. Examples are (but not limited to): getting a forecast up and running, developing teaching materials, finalizing tutorials, refining or creating tools, analyzing forecasts for a manuscript, and/or developing visualizations.
         2. The event will also include a poster session for attendees to present their research.
      ii. Travel funds within the US are available
      iii. Space is limited to 50 people.
      iv. The application to attend is due on Feb 1. But the nominations of project ideas are not due at that time. So this group will have more time to think about specific project ideas to propose.
   b. Updated Forecasting Standards Manuscript pre-print is available https://ecoevorxiv.org/repository/view/4054/ The manuscript was submitted to Ecospheres on Jan 17.
      i. The manuscript "summarizes the open community conventions developed by the Ecological Forecasting Initiative (EFI) for the common formatting and archiving of ecological forecasts and the metadata associated with these forecasts. Such open standards are intended to promote interoperability and facilitate forecast communication, distribution, validation, and synthesis."
      ii. The Community Standards R package associated with the manuscript is archived on Zenodo: https://zenodo.org/record/7494824#.Y8bhLnbMJD8
   c. Leah shared her new VecTraits resource, a database that provides functional traits for disease vectors
      i. https://vectorbyte.crc.nd.edu/vectraits-explorer
      ii. https://www.vectorbyte.org/blog/introducing-vectraits

2. Previous Project Ideas
   a. Survey to assess people's interests on the following projects/activities
   b. Data Ingest, Cleaning, and Management Task View

     c. Workshop Proposal to identify gaps in CI/methods for forecasting and put forth a proposal for a workshop to bring people together to work to fill those gaps
        i. Google doc to collectively develop ideas
     d. Develop forecast visualizations for the NEON Challenge
     e. Discuss papers that cover existing cyberinfrastructure or stats methods
     f. Organize, present, or view presentations about topics in cyberinfrastructure or a stats method
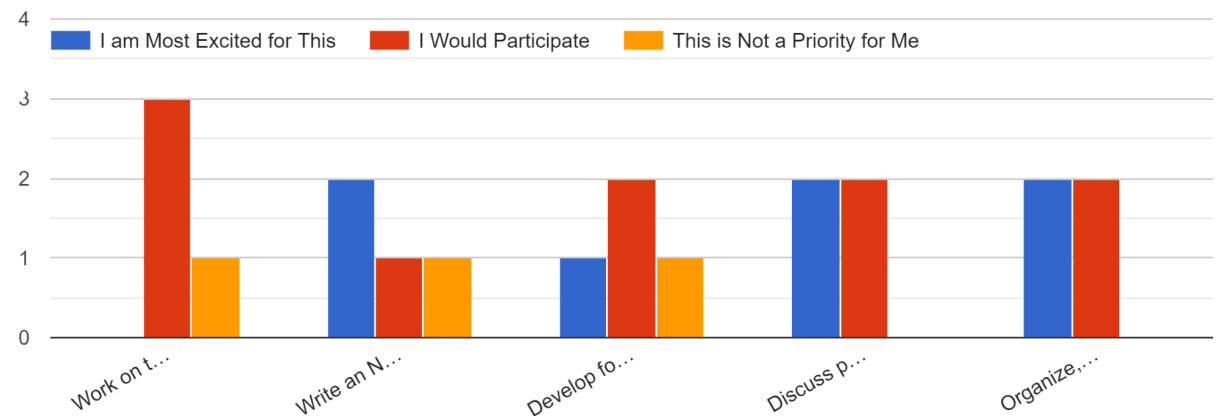
3. Results from the Survey
     a. 4 responses

Rank your personal goals for participating in the Cyberinfrastructure and Statistical Methods Working Group?
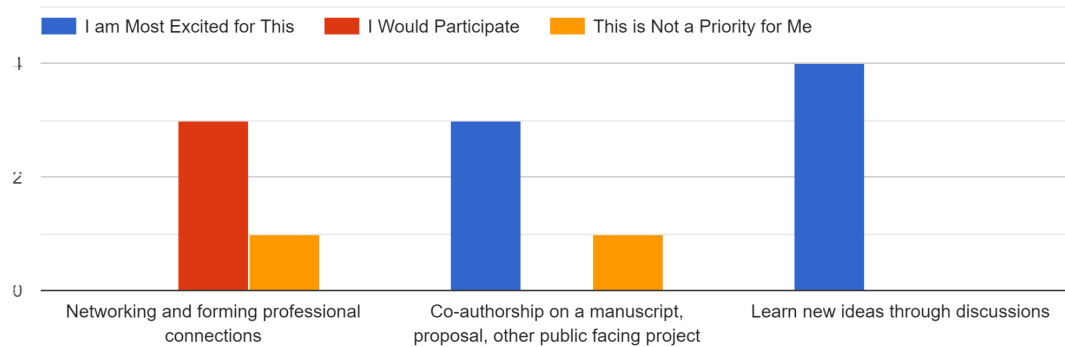
Other goals - make connections between EFI and NASA, improve forecasting expertise

The following are Activities listed during recent CI/Methods Calls. Rank what you are most interested in.

4. 1 - Work on Data Ingest Task View

2 - Write NSF workshop proposal
3 - Develop forecast visualizations for NEON Challenge
4 - Discuss CI/stats methods papers related to forecasting
5 - Organize, present, view presentations on CI/stats methods related to forecasting
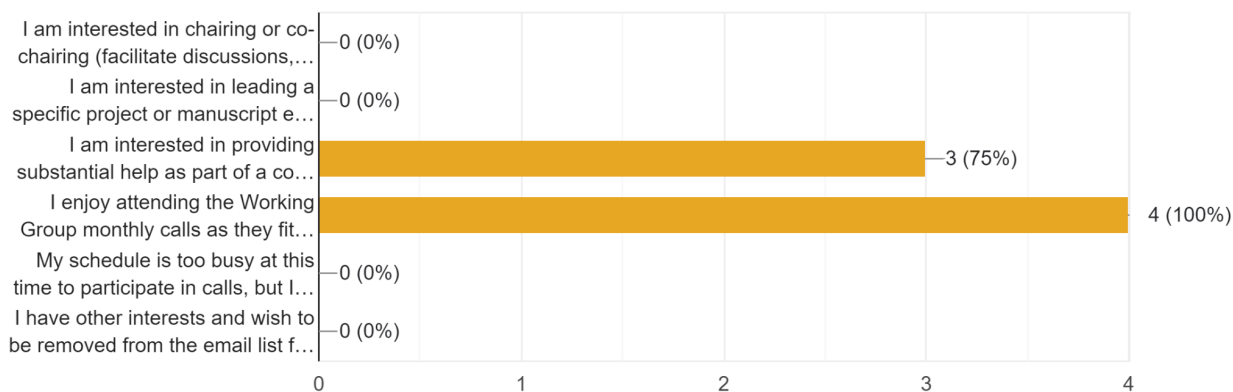
Other activities suggested:
Maybe explore value of a NASA UNBOUND workshop venture, depending on where nsf one goes
Learning more about current methods that are happening for running forecasts in production type environments with stakeholders.

We need people to provide leadership for the calls and for projects. What is your availability?
4 responses

Notes from the Discussion on Today's Call
1. In regards to Unconference - could have this group provide a list of wants and needs to help people attending the Unconference narrow down how to solve a problem instead of people not feeling included because they don't have ideas to share but instead are interested on working on projects.
2. Stats Seminar has been very successful. If we did - do we want to focus on tools ecologists use that is distinct from statistics. Could be getting data ready for stats
   a. Disconnect between participation in the Stats series and the participation in this stats working group is large. How can we more actively help people cross over?
   b. Instead of creating another seminar series on CI side, think about how can we leverage the Stat series to bring people into other working group activities?
   c. What can we announce at the next stats series to get people involved
   d. There is a lot of demand to learn things, but it is less obvious on how people can contribute. People present, but there is not incentive for participants to do

anything. How to lower the bars of entry for participation. The value added is much clearer on the learning side

    e. Could there be deeper dives for projects?

    f. If we do a webinar on something related to the forecasting challenge and then have a follow-up. You just did this webinar. Here is the opportunity to break into small groups and apply this to a NEON data stream.

    g. Come to a workshop where you come to learn to forecast and in the process you contribute a forecast has been the best way to get people to submit forecasts

    h. Have lots of submissions to the Aquatics Challenge because of the workshops. Could put a couple of workshops on the books to be able to point people to.

    i. Think about if it is possible to put effort into other Challenge data streams to set it up to be similar to the Aquatics Challenge.

    j. For the aquatics the model to convert air temp to water temp is a simple model to get people's mind around. With phenology it is harder to come up with something similar - having a slightly more complex model than the null models.

    k. Want a simple conceptual model to riff off of. What is the simplest model that uses the weather forecasts that could be inserted into the template that is developed for the Aquatics Challenge

        i. Here is the Aquatics workshop material: https://github.com/OlssonF/NEON-forecast-challenge-workshop

    l. Carl created a tutorial website based off the Aquatics

    m. Think about the axes where we can lower the bar to peak people's interests

    n. Is there something that can be done in a spatial context

    o. 1 hour workshop is a good way to have people set time aside to learn that may help people engage that complements online material that people can look through

3. Reading group - the semester that there was a reading group on spatial temporal analyses.

    a. The thing people are most excited about are learning so workshops or reading groups are effective

    b. Could repeat the spatio-temporal book previously discussed or pick another book with hands on activities

4. Workshop proposal

    a. Jake has been trying to gauge interest at different federal agencies to participate in a workshop around CI/methods. There is a lot of interest.

    b. The question is would there be agencies/groups that would adopt what comes out of the proposal (e.g., white paper, recommendations)

        i. The timing at USGS is good for this

        ii. There is a new project applying models at larger scales. This year the group is planning on how to do that. Looking for recommendations and practices

        iii. Timing from USGS side - would be good to do 6 months from now or before end of fiscal year in Oct

              1. Now-ish funding asks would be best for aligning with NASA fiscal year spending
-     iv. From NASA Biodiversity and Conservation this would be a good time to work on this
- c. Mike reached out to Peter McCartney at NSF about a workshop and he liked the idea of having a white paper as an output
    - i. A good outcome of the workshop is a high level design document that would allow people to move forward in an interoperable way so NSF could implement X, USGS could implement Y, groups could implement things that would be interoperable
    - ii. Peter would like contributions from ecology and CI sides. NSF XSEDE - people who know the industry standard and the cutting edge info for CI
- d. Think we could get input/funding from multiple agencies.
- e. What would be useful for people ranging from (for example) agency people to assistant professors
    - i. 3 things to have at the end of the workshop
        1. Design principles
        2. References and implementation (examples of workflows/models people could find to not need to reinvent the wheel)
        3. Validation or validator to check
            - a. Standards is a good starting point for what to check for.
    - ii. Piece together our experiences with a lot of other voices in the room. People coming to the meeting will know the scope - automated tasks with predictable inputs/outputs that can handle up to gridded global forecasts
    - iii. Get input from the different participants on what the different use cases are in order to triangulate what has been done
- f. USGS has things put in place but not sure if it is the correct way to do it or to find out what others are doing differently
- g. Depending on timing and if we want to capture CI folx in US fed agencies, ESIP summer conference is in July in person this year. https://www.esipfed.org/meetings
- h. [NASA ESTO](#) [AIST](#) is interested in beyond ecology in identifying and removing inefficiencies in workflows and have decent budgets.
    - i. Also may be worth keeping their funding opps on EFI radar (e.g., here is an ESTO-funded [biodiversity proj](#))
- i. Next steps
    - i. Write 1 or 2 pager and shop it around?
    - ii. Develop a Google doc with ideas to then shop around to Agencies
    - iii. Short list of Agencies to shop it around to
        1. Start with USGS and NASA
    - iv. Primary funder NSF?
        1. Depends on how much we are asking and what people require
        2. By asking USGS/NASA to contribute funds then they will be more likely to contribute to the outcomes and buy in

3. If we ask NSF, USGS, NASA each for $20K then will have plenty of resources for workshop

    v.   NSF may say to use RCN funds. But we could have a unique ask. RCN is about NEON ecology. The purpose of the workshop is to broaden it to other US agencies.

    vi.   Jake has time in about 2 weeks to work on outline and send out. Jessica is also able to help with it.

    vii.   Mike will be point person to go back to NSF to see what they

    viii.   Chris thinks people at USDA would be interested. USDA wants to bring all USDA data and models together so it is searchable. Chris will check in with that group

    ix.   Reusable models is part of the forecasting workflow, but you reuse models for many different reasons.

    x.   People from NOAA should be invited

    xi.   Also think about people from technology companies. People who know the ins/outs of Amazon services. But it has to the be the right person.

    xii.   Google does flood forecasting so there are examples of Google forecasting

    xiii.   Mike has connection to IBM

5. What can we seed for Unconference projects? We envision people submitting pitches and viewing pitches. So if we can provide examples of pitches to add to the list of projects that would be good
   a. Spatial forecast (Carl) - have been forecasting netcdf/csv of timeseries, but have nothing spatial. Would like to expand to rasters/other spatial forecasts. Example: MODIS leaf area forecast over the next year
   b. Post disturbance recovery can be interesting to predict. If you predict LAI annually it is most interesting to predict in places that have recently been hit by pests or fires. Predicting the disturbance is hard, but predicting the post disturbance recovery could be easier
   c. Forest insect - have gypsy moth trap data for past 20 years. Could do population level gypsy moth forecasts with spatial forecasts and disturbance recovery. Chris has the data and could have it cleaned up in time for the workshop. WI to NC spatial extent. Could also ignore the space and focus on the population forecasts as well
   d. Reviving the Tick Challenge
      i. Need machine learning techniques to count ticks for tick counts for NEON
      ii. Product example for the Unconference related to this would be could be a pathway to getting this with specifics - we need to contact this person
   e. Expanding tutorials to the other themes
6. Next call
   a. Follow up on workshop proposal
   b. Continue to brainstorm education activities - short term ideas for Unconference. More long term ideas for educational opportunities related to CI/methods for the broader EFI community