

August 22, 2023 Joint Methods & CI Working Group Call

Attendees: Joel Mota, Carl Boettiger, Jody Peters, Quinn Thomas, Brittany Barker, Leah Johnson

Regrets: Jessica Burnett, Jake Zwart

Agenda/Notes:

1. CI Workshop Proposal Update (Jake, Jessica, Chris); meeting later today to discuss next steps
 - 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. Meeting 1 will be hosted at NERACOOOS
 - i. Have funds available from NOAA and USGS. Submitted proposal to NASA for additional funds
 - d. Other funding options to keep in mind perhaps for Meeting 2
 - 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.
2. Standards Manuscript Update - waiting to hear from Mike about how he is doing with getting the proofs sent back in so the manuscript can be published.
3. Update from Unconference Activities and Next Steps
 - a. New blog post summarizing dashboard updates from one of the Unconference project groups:
<https://ecoforecast.org/reenvisioning-efi-rcn-neon-forecast-challenge-dashboard-visualization/>
 - b. Anyone can provide feedback about the dashboard. For now use the Slack or ping the GitHub issues. Or reach out directly.
 - c. The dashboard is created by Quarto, it is a static page that gets developed daily
 - d. dashboard lives at: <https://projects.ecoforecast.org/neon4cast-dashboard/>

- e. catalog draft lives at:
<https://radiantearth.github.io/stac-browser/#/external/raw.githubusercontent.com/eco4cast/neon4cast-catalog/main/stac/catalog.json?.language=en>

4. Forecasting Challenge CI Updates

- a. STAC [prototype](#) for making the Challenge forecasts searchable
- b. There are a lot of people interested in running their own Challenge in a different area - e.g., USGS forest chl-a for rivers in the US. Terrestrial network in Australia. Virginia Tech local reservoirs for drinking water. S. African chapter challenge around their data which is similar to NEON's data.
- c. The goal is to develop the CI to support different types of Challenges
- d. Does it make sense to have this working group lead that working group or set up a new one to give people a place to ask questions about running or leading a Challenge?
 - i. Will depend on the other groups and if they want to show up.
 - ii. Right now individuals email us and Jody/Quinn email back, but could
 - iii. Could also try it as a Slack channel since there are group/people across multiple time zones
 - iv. How much that is asked is related to CI and what is asked that is related to data, etc?
 - 1. Both
 - 2. Most recent group from ad hoc group from Canada - their questions have been we have this data stream, is it well suited for a challenge and have talked about data latency (short term data is easy) vs questions that are interested in more long term carbon storage
 - 3. Outputs and standards formats are part of the CI - being able to reuse scoring and dashboard involves knowing the standard
 - 4. Also introducing the concepts - target files, what is a score, etc
- v. Are there examples of how the Challenge has been used in the classroom?
 - 1. Yes. See this:
<https://projects.ecoforecast.org/neon4cast-docs/UseInCourse.html>
 - 2. Also NEON wrote blog post about it:
<https://www.neonscience.org/impact/observatory-blog/efi-neon-for-ecasting-challenge-classroom>
 - 3. This also leads to the question of whether people should join the Challenge working group who are interested in using it in education or if that would be separate.
 - 4. There are also workshop materials (this is the one at ESA):
https://github.com/OlssonF/ESA2023_neon4cast_workshop
 - a. Good example of how to get a forecast going
 - 5. Beetle example:
<https://ecoforecast.org/collaborative-innovation-and-skill-building-a>

[t-the-2023-unconference-empowering-ecological-forecasting-with-neon-data/#ground-beetles](https://github.com/OlssonF/cefi_shortcourse_students_2023/tree/main/CaseStudy_WaterQuality)

6. Water quality forecasting example with a hierarchical Bayes state space model:

https://github.com/OlssonF/cefi_shortcourse_students_2023/tree/main/CaseStudy_WaterQuality
 7. Suggest to Education working group the need for pulling together educational materials more broadly related to the Forecasting Challenge
- e. Chris Jones update from Unconference project - Goal - is it possible and how to do uncertainty partitioning for forecasting and adapting it for different statistical models.
- i. Chris and Noam are still working on the ML component of this project and trying to decompose the uncertainty using tidymodels. Used random forest. At the time it seemed like the easiest option, but not sure if it is.
 - ii. Mike, Eli, Amanda were working on the statistical model
 - iii. Goal was to determine if it was possible or if it is not possible. If possible, then how to do it
5. Next Steps - What does the group want to work on? Continue to hold joint CI/Methods calls?
- a. Scheduling monthly calls
 - b. Brittany - protocols for standardizing methods in general. Data ingest and data availability could influence the way the methods are chosen and uncertainty. Hve resource that people can visit when they are starting modeling - a resource they can look at to see what most
 - c. Quinn - Think about a style of a thing we want to work on that is reflective of the bandwidth. What is the style of work we want to do?
 - i. Tutorials?
 - ii. Mini-sprints to work on projects that we want to work on collaboratively?
 - iii. Brainstorming and sharing of best practices
 - d. Joel - built 1 model for his MS research. But interested in working collaboratively on projects to learn and grow.
 - i. Then decide what the group wants to work on and what does the group have bandwidth to work on?
 1. Tool for community or Rmarkdown tutorial to show the community something
 2. Example of tutorial - book: Forecasting Principles and Practice (<https://otexts.com/fpp3/>). They use business datasets throughout the book. There is a whole opportunity to use the NEON Forecasting Challenge to walk through the different techniques used in the book so people can translate business examples into ecology example

3. Zooming out - this book is great but is most related to the Methods, but has no relevance to CI.
 4. Theory working group is also working on some methods projects.
 5. Brittany's example
- e. Chris - I have a forecasting, now how do I automate it?
- i. On some levels we have that available as a solution. So now we are at a good time to think about what is next. We have nothing close to the book.
 - ii. Right now we only have the documentation. conceptualize what are the boxes. Not everyone interested in automating a forecast are interested in scoring 50 teams every day.
 - iii. Provide friendly tutorial information on how to.
 - iv. If you want to automate an aquatics forecast we have a good tutorial. But if you want to apply it to something else then you are more on your own. There is nothing outside the context of that example and that fits squarely into CI.
 - v. People coming to that working group will have questions about automation and reproducibility. Going to methods is more about model development.
 - vi. But it could be an issue if it splits us to thin and we don't have enough people to go to both groups
 - vii. Are there people who would go to methods call if it was focused on the model generation?
 1. Think so and if it was scoped more broadly. Let's talk about ML in forecasting and time series forecasting. The goal is to have the group be a tutorial on how to do it and where it fails and succeeds
 2. Could do the same on the infrastructure side - could lean into the automation side of things. Preparing visualizations or storage or data integration.
 3. Working group could also oversee a docker container that the group revisits and sees if there needs to be more software or updates need to be made for use in forecast generation.
 - a. Here is a stable EFI forecasting container - what needs to be in it. What makes it too heavy?
 - b. Interesting, especially if you have a spatial forecast which requires a different set of tools than others - have 2 dockers. One for spatial and one for aspatial.
 - c. If we want to target the spatial stuff, then Chris is happy to lead.
 - viii. How much is there for rounding out the edges for the spatial example from the Unconference?
 1. Think there is enough to get started to generate a forecast, the target, and score them using the infrastructure that we already have.

2. Carl is willing to run with the spatial example which touches on a bunch of things we have talked about.
 3. For CI - the spatial has a lot of data size aspects that are difficult for automation and pipelining.
 4. Brittany is also very interested in the spatial example and it does touch on Methods related to dealing with large data.
 5. Joel is also very interested
 - ix. Jody to send out a launch announcement.
 - x. Reach out to Emma, John, and Dave from the Unconference as well. They are all still invested in it.
6. The following are topics that have been discussed previously
- a. Brainstorm educational activities that could be developed/promoted by this group
 - i. Longer term ideas for educational opportunities related to CI/methods for the broader EFI community
 - ii. Stats seminar-like presentations or how to use the Stats webinars to bring people into EFI activities
 - iii. Reading group?
 - b. Data Ingest, Cleaning, and Management Task View
 - c. Discuss papers that cover existing cyberinfrastructure or stats methods
 - d. Organize, present, or view presentations about topics in cyberinfrastructure or a stats method