

## June 1, 2020 Joint Methods & CI Working Group Call

Attendees: Alissa Brown, Chris Jones, Thilina Surasinghe, Rob Kooper, Volodymyr Trotsiuk, Abigail Lewis, Matthew Helmus, Mike Dietze, Jake Zwart, Jody Peters, Yang Song, Bruce Wilson, Ben Toh, Quinn Thomas

Agenda:

### 1. Introductions

a. Jake's presentation of the CI and Methods group for the RCN:

[https://youtu.be/CB\\_ZUDIerpM](https://youtu.be/CB_ZUDIerpM)

b. Introductions:

- i. Mike - EFI Director at BU. Lab does lots of eco forecasting projects and has done a number of CI/Methods project (e.g., PEcAn - CI support for terrestrial modeling)
- ii. Jody: EFI project manager; works with Jason McLachlan at U. of Notre Dame
- iii. Jake - postdoc with USGS in water mission area. Forecasting stream/lake temp and connecting process based models with deep learning models. Got involved with EFI through summer course a couple of years ago.
- iv. Abby - PhD student with Cayelan Carey at Virginia Tech. Forecasting with lakes. Interested in learning about CI and Methods
- v. Chris - research prof at NC State - forecasting pest and pathogen spread.
- vi. Matt - assistant prof at Temple University/ forecasting patterns. Work with Chris modeling the spread of forest pests and modeling spread of species across islands. How to communicate forecasts to stakeholders
- vii. Alissa - interest in joining this group is learning about CI. Postdoc at Morton Arboretum response of trees - genetic data, pollen data, modern occurrence data
- viii. Rob - work with Mike for last 11 years. Background - computer science. Interest in CI
- ix. Thilina - Bridgewater State University in MA. New to forecasting. Interested in looking at community changes in restored wetlands
- x. Bruce - Oak Ridge DAAC. Main interest in the group is as provider of data and has been doing CI to support science for last 20 years
- xi. Volodymyr - research scientist at Zurich. Tree ring ecology background and now in forest growth models. Long term ecosystem research - similar to NEON but in Europe
- xii. Ben - PhD at U of Florida. Marine biology/ecology background. Currently focused on using statistic models and machine learning on malaria.
- xiii. Yang - Assistant prof at U of AZ. research - how to simulate microbiotic community and vegetation and their role in water and land/atmosphere connections. Interested in data integration and the model approach
- xiv. Quinn - Associate prof at Virginia Tech. Forecast forests and water quality. Leading EFI-RCN.

2. Things to come back to after the RCN:
  - a. When we work on archiving platforms - we need to have a follow-up call with CyVerse
    - i. Figure out pros and cons for different platforms. Make recommendations for the RCN
    - ii. Requirements -
      1. need to push forecasts automatically on a daily basis
      2. Searchability of metadata. Want the metadata to be searchable
  - b. Need CI plan for how we will move forward as a community
    - i. Good ideas were thrown out during RCN Breakouts for high-level CI needs
      1. CI Breakout 1 [link removed]
      2. CI Breakout 2 [link removed]
      3. CI Breakout 3 [link removed]
      4. CI Breakout 4 [link removed]
    - ii. Will run forecasts in multiple iterations. Will have a plan for 2021 forecasts which may be different than the plans for 2022, 2023, etc
3. EFI Task View Plans
  - a. [Task View Post 1](#) is done
    - i. Led by Jake Zwart and Alexey Shikalomanov
    - ii. Reproducible Workflows for Forecasting in general. Included 10 different categories with a paragraph about the value of the task and highlighted a couple of the tools for that task.
    - iii. Living document. Metadata section could be updated further moving forward and other tools can always be added
  - b. What do we want to do for future Task Views:
    - i. Uncertainty Quantification & Propagation, Modeling & Stats
    - ii. Data Ingest, Cleaning, Management
    - iii. Visualization/Decision Support Tools, User Interface
  - c. The Task Views serve multiple purposes
    - i. The idea is that these Task Views are similar to a Cran R list of resources
    - ii. Here are the tools people in the community are using
    - iii. Resources for folks in the community
    - iv. Gap analysis - if we identify and prioritize the tools the community use for different tasks in forecasting, then we can also see where we are missing tools and where no one is happy with the current tools that are available.
4. Other needs we discussed before the RCN
  - a. Put together short vignettes for common needs (downloading NEON data, downloading met data, etc). Use short vignettes instead of a full blog post on multiple tools that will take longer.
    - i. Even having a workflow that runs a null model will be useful as others new-to-forecasting could build off of the workflow
5. We have an ecoforecast GitHub repo: <https://github.com/eco4cast>

- a. Comment to follow up on from the May 8 call - Add ReadMe which lists tools we want to add
- 6. Timeline for RCN Forecast
  - a. Not sure when the first ones will come online. Shooting for a start within the next calendar year. 6 months - 1 year from now
  - b. Will start with 5 forecasting groups
    - i. Phenology (from population and terrestrial groups)
    - ii. Water and carbon land-atmosphere overlap - flux tower
    - iii. Ticks
    - iv. Aquatic Instruments
    - v. Populations/Communities - some topic that still needs to be defined
  - c. Polls will go out to schedule these groups to talk starting by the end of June
  - d. From Rob Kooper: Spoke with XSEDE director and he let me know that there is the ability to leverage of XSEDE resources for the challenges proposed.
- 7. Example: Chl a - comes off the sensor pretty much ready to be forecasted (as compared to eddy covariance which has to be processed)
  - a. Do we want to test run the collaborative protocols?
  - b. Or does one person take the lead and others chime in?
  - c. Use GitHub so someone can get initial files committed and others can submit pull requests to improve
  - d. Make it modularized - one person works on ingest, one person works on pushing to archives, etc
  - e. Set aside time on next call to come up with high-level outline. What are the major modules that would go into the basic workflow. How can we organize? And who is interested in tackling the different parts?
  - f. The modularity is nice because it can be applied to other data products.
  - g. Homework for next call - create a Google doc for folks to collaboratively outline the major modules, prototype workflow to help with forecasting challenge**
  - h. Large scale data products needed for forecasts
    - i. RCN Steering Committee talking to NOAA about the met data
    - ii. Partners Working Group - facilitate identification of common data products used by ecoforecasting community. This is synergistic with data Ingest task view (that focuses on the tools) while this would focus on weather forecasts, DEM layers, satellite, GIS files. Finding a balance between self-promoting my data for my site and thinking about it from a higher level - what are the common data resources in different categories that the community is using. Also leaning towards datasets that are continually updated (although things like DEMs, soils maps are not necessarily going to constantly be updated or change between versions)
- 8. Matt - interested in focusing on the Interface between scientists and people who use forecasts (users)
  - a. This is partly the domain of the Partners group. One of their goals is co-production.

- b. Social Science group looks at how forecasts should be used to make decisions compared to how they are actually used.
  - c. CI/Methods Task View on Visualization - will need to have calls that bring the Social Science/CI/Methods groups together to look at best practices/tools for this task.
9. Next Task View
- a. Modeling Stats - Abby, Ben, Alissa happy to help
  - b. Visualization - Abby, Ben, Alissa happy to help, Chris happy to co-lead with people more familiar with R Shiny and tableau
  - c. Data ingest - does that include consolidating across data sources?
    - i. Data harmonization - how you combine different data from the CI perspective (2 compatible datasets generated by 2 teams). This falls within Data ingest
    - ii. Statistical data integration/statistical data fusion - falls within Stats task view. Mike expects that statistical data fusion will show up as a gap
  - d. **Homework - Drop bullet points in to this doc over the next month:**  
Uncertainty Quantification & Propagation, Modeling & Stats [link removed]