August 24, 2020 Joint Methods & CI Working Group Call

Attendees: Helen Scott, Abby Lewis, Ash Griffin, Kenton McHenry, Quinn Thomas, Ben Toh, Jake Zwart, Bryce Mecum, Jody Peters

Agenda/Notes:

1. Updates on EFI Task Views
    a. Use Task View 1 on Reproducible Workflows as a guide
    b. Uncertainty Quantification & Propagation, Modeling & Stats
        i. review the restructure/outline that Abby, Ben, Ash worked on over the past month
        ii. Assign people to fill in text for different sections
        iii. Want advice on tools for different sections and suggestions and examples that can be added
        iv. Currently have it structured with Intro with example of something you want to forecast, have drivers, have a model that uses the drivers, and have uncertainty
        v. Have a section on models and uncertainty in general for people not familiar with it. This is a good place to talk about places of uncertainty. How to get met uncertainty, e.g., ensemble weather
        vi. Tools and technique section divided by stats model, bayes model, machine learning, agent based models
            1. For each model have description of model, a section for tools and other resources and uncertainty as it relates to each model
        vii. Adding in something about Data Assimilation
        viii. Synergy is with the Standards for archiving forecasts
            1. There is a section in the Standards is about defining uncertainty
            2. There is a debate within the Standards group about whether it is required or optional
            3. If required needs high level of understanding
            4. When people are filling out the Standards and get to the Uncertainty section (2.1.2), if people are confused we can point them to this Task View
        ix. Stats Models Section of Task View
            1. **If anyone from this group knows of papers that are useful for describing the kind of models or use cases for the model examples (ARIMA, Species Distribution Model, etc)**
        x. Bayesian framework of Task View
            1. Talk about similar models as stats section, but talk about priors, etc
        xi. Machine Learning
            1. Design trees, neural networks, etc

2. **Would like input from this group on papers using machine learning for forecasting, especially if there is any input about uncertainty with machine learning**

xii. Deterministic Model

xiii. Agent Based Model
   1. Ben is working with this model for his disease research, but isn't sure of examples for forecasting.
   2. **Would like input from the group about any papers that use agent based models for forecasting**

xiv. How much to outsource to other documents? Could be a long task view if it includes everything right now.
   1. First task view tried to keep each subsection to a page each. Was around 10 pages in total
   2. Don't need to do much more description for each of the models. But for each of them there are tons of resources that we can point people to.
   3. First section - what is uncertainty and how to choose a model can be pretty short
   4. If we define how the forecasting standards is relevant to each model, then that could take up space
   5. Have headers for each model with a link that takes you to the list of resources.
   6. It would be nice to have a little carrot that brings up a dropdown with the list of resources, but Jody isn't sure how to do that with wordpres with our theme. Ash will look into this and see if he can find information.
   7. Jake can add example of uncertainty/forecasting for machine learning
   8. Working example - chlorophyll a forecasting. **If anyone knows of chla forecasting examples for any of the types of models**. Offer a workign citation with an example. If working citation emphasized the same system. Here's bayes forecasting section and here's an example of chla forecasting using bayes.
      a. A really great example is key (if it is about chla then it would be good).
      b. Code example or paper example? Hypothetical example
         i. Paper and if the paper has code then even better
   9. Next tasks/timeline
      a. Ben, Ash, Abby to meet together before next call to fill in the text
      b. People in the group should add examples
      c. Abby, Ben, Ash to flag where we need to outsource

c. Data Ingest, Cleaning, Management

i. Matt H., Jake Zwart happy to help, Chris Jones happy to help, Bruce happy to help
ii. Any progress on this in the last month?
1. No progress yet

d. Visualization/Decision Support Tools, User Interface
i. Abby, Ben, Alissa happy to help, Chris happy to co-lead with people more familiar with R Shiny and tableau.
ii. Social Science group is meeting tomorrow and will walk through this as a group to add resources

2. Forecasting Workflow Updates
a. Google doc outlining major tasks here:
b. Update on GitHub phen-null repo
i. Has Kathryn/Dietze lab moved her phenology forecast to the new repo.
1. Mike hasn't been able to get to this yet, but it is on his list
ii. Any needs in terms of generalizing to EFI standards?
c. The beetle and phenology RCN groups will have workflow examples to start with to provide a template that the other RCN groups (aquatics, carbon/water flux, ticks) can use
d. Will there be a shared tool for getting driver data? Will it be produced with the first examples?
i. Quinn is tasked with creating a container to have the NOAA met. It will probably on CyVerse. It will run every day and put the data there. This will be provided by the RCN. If we get this working, then will work on longer range forecasts from North American Model Ensemble
ii. If there ar eother covariates the groups need, they will be responsible for them.
iii. Directory will be set up with: Product, site, date, and a bunch of ensemble members in that. 4 times a day for every NEON site.
iv. OakRidge DAAC have been subsetting data and the RCN will point folks to that. It isn't as real time - if you want LAI with MODIS from NEON sites you can get that so you don't have to process the data yourself.

3. Archiving platforms - see the table on pg 23 of the v2 Forecasting Standards.
a. ID - if you produce a DOI for every forecast it blows up. Can you create a subID
b. Is there a preferred archiving platform. Discussion is ongoing

4. Reminder that we have this EFI Software Needs GitHub repo. Use the issue tracker to pose software, tools, coding needs.
a. Are there tools for the Challenge and in general for the forecasting community. Good for hackathons.
b. Would each new tool get its own GitHub repo? TBD.
i. NEON has NEON Utilities which is a large package that downloads lots of things (one thing you might need and lots that you might not need)

ii. Do we want an EFI package that pulls in the drivers, data that you need
iii. Expect to have tools related to each other are in the same repo, but a bunch of different repos