

November 17, 2020 Joint Methods & CI Working Group Call

Attendees: Helen Scott, Alexey Shiklomanov, Kenton McHenry, Quinn Thomas, Ben Toh, Rob Kooper, Leah Johnson, Carl Boettiger

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. Updates from Ben
 - ii. Have reorganized to be in more the Cran framework listing packages
 - iii. If anyone has suggestions for tools to add, it would be great to have you add them
 - iv. Jody and Ben to check in with Istem about getting text for the black boxes sections
 - v. Processed based section - there are a billion processed based models. Will this section introduce people to them?
 1. Community Land Model - has executable that you need to know how to run. This is out of the scope of this. In scope is the idea could talk about how to solve the problem.
 2. Could talk about ways to call executables from python/R
 - a. There are lots of tools for this. In R has a number of packages to solve ODEs (Leah)
 - b. How to call in executable
 - c. Things in netlogo to build agent based models
 - d. People will need to design their own (except for example something like SIRE models)
 - e. Gillespie method and chemical kinetics (in R GillespieSSA)
 - f. There are things available to help people, but don't expect people's model will ever be off the shelf
 3. Have package that fit ODE, and others that fit mechanistic models
 4. Assume people will not be building these large models. How do you do stats when you are handed a FORTRAN model and what do you do with it?
 5. Tools for how to run these models/tools in R. How to get data in/out of them.
 6. What tools can or can't you use with them.
 7. Any black box model - can't use JAGS or Standard BUGS
 8. Things you can do with emulating models (Istem's experience)
 9. Think of it in this way: "I've been handed this big statistical model - what are some of the tools I can use"
 - vi. Eco forecasting involves data interacting with models. This is where R/Python come in. Scope of this Task View is how to use R or Python

- vii. Uncertainty section
 1. Do we need this section?
 2. If the Task Views are for pointing people to resources. Do we need this section since this is more tutorial and in many cases expect that people will simulate or code their own uncertainty analyses
 3. R tools for helping to do bootstrapping and ensemble analyses when you don't propagate the uncertainty itself
 4. Not many packages that do uncertainty
 5. Boot package - does some bootstrap but is limited
 6. Help people guide people through Uncertainty is important, but it may not be a Task View
 7. Next thing you want to do is to do some uncertainty quantification. Here are some approaches you could take and then could link to tutorials.
 8. Provide an overview of Uncertainty and the Types of Uncertainty but since there are not many off the shelf resources to quantify uncertainty, then just point people to the tutorials
 - viii. Data Assimilation section - Jake has provided a link to some tools
 - c. Visualization/Decision Support Tools, User Interface
 - i. Placeholder while Whitney takes her prelims
 - d. Data Ingest, Cleaning, Management
 - i. Placeholder until we are further along with the other Task Views or have an identified leader for this
2. Forecast Data Visualization Shiny App Update
 - a. Bruce Wilson and the ORNL DAAC's experience moving from RShiny to Python Flask
 - b. Bruce wasn't on the call. So we'll add this to the Agenda for the next call.
 3. [NEON Ecological Forecast Challenge](#) CI Update
 - a. At a stable point.
 - b. NOAA forecasts are still downloading
 - c. Waiting for last bits of information from the design teams
 - d. Took perspective of taking a generics R container that doesn't come with special packages, but instead have a generic set of packages
 - e. If design teams have specific packages they need would need to contact Quinn/Carl to get packages added
 - f. OR currently are launch cron job that puts the R script through the container and then the Cron job needs the packages to run the R script
 - g. Carl playing around with an R package that is the update to packrat. It works well except it is hard to upgrade. It is good for frozen versions. But in the development stage it gets in the way. However, if it is taken away, then don't know what packages individual teams need

- h. Any ideas of solutions?
 - i. Other case is for people using non-R script
 - j. Could have a docker file that extends docker image for people with special CI needs
 - k. Could have vanilla container that runs vanilla models for each theme. But there may be technique needs that specific teams will need.
 - l. Rob did the docker file that extends the docker image.
 - i. Someone wanted python 3.7, someone else wanted python 3.9, etc
 - ii. Rob ended up building 20 base images for different python versions. He ended up scrapping
 - iii. Could start with a container with more python built in. Versioning with python is more severe compared to R.
 - iv. Give people a starting container that will cover 90% of use cases and anyone that doesn't want that, then it is up to them to extend it and we can work with them on it
 - m. Have general use container that we want someone to give us R script and we put it through it. How do we do it? Give the file that has both R and Python or at minimum give us an R description file. Then 90% of the time can install packages
 - n. Right now Quinn/Carl installing packages as they go
 - o. Now can load large amount of soil data over multiple sites over multiple year without crashing your computer
 - p. At some point need to work on the scoring pipeline nailed down and automated. Computing the scores.
 - q. Need scripts that move the submission files to the forecast location
 - r. Do some kind of validation on the metadata. If it doesn't work, an error message goes to the teams.
 - s. Also want that the script that does the sorting it decides which bucket it goes into (tick, beetles, etc)
 - t. Looked over Quinn's diagram of the Forecast Challenge CI
 - i. Will want to add this diagram to the website
 - u. Contributing teams only will need to know how click on an icon and download data from MINIO. Submissions are also as easy as clicking a button. Or could be automated.
 - v. At this round we will get the metadata, but won't be able to automatically reproduce the code yet.
 - w. Would it make sense to say here is a snippet of code to download/upload files. The design teams will have that for the null models so that people can use that as an example
 - x. Do we have snippet of code for how to download data on the website? Yes. We have R but don't have python right now
4. Forecasting Workflow Updates
- a. We discussed this on the October call so there may not be anything else to go over, but Jody is leaving it in as a placeholder in case there is more to discuss

- b. We had a short amount of time to look at the Workflow on this call and added a few more places where the Challenge will cover some of the Workflow steps.
- c. On the next call we can quickly go through the rest of the steps for one final review.