

## March 27, 2020 Joint Methods & CI Working Group Call

Attendees: Alexey Shiklomanov, Mike Dietze, Quinn Thomas, Jake Zwart, Chris Jones, Yiluan Song, Carl Boettiger, Kenton McHenry, Jody Peters, Rob Kooper

### Agenda:

1. Update on EFI Task View Intro and Workflow/Pipeline Overview (Jake)
2. Methods/CI Working Group text on [Website](#) - do we want to update it?
  - a. See individual [Methods](#) and [CI](#) pages
3. Follow up on discussion from last method call about forming a subgroup within intersection between Methods/CI interested in spec-ing out what a general met workflow (or expand to grid data) looks like.

### Notes:

1. The Overview document is a living document that will be posted to the EFI website as a News/Blog post and with a link to the CI/Methods pages. Jake needs to look through the comments. But getting closer to putting on a website
2. The overall Task View has 9 areas that have been grouped into 3-4 areas that can be put up as 3-4 blog posts on the EFI website.
3. The Overview that Jake has taken the lead which will focus on the 1st area.
4. Kira has worked with the Social Science has helped create a [blog post on visualizations](#) that we will cross-reference when we get to the visualization area from the Task View
5. We want our Task View to be similar to Cran's task view which helps people find packages that are useful for certain topics (e.g., Useful tools for Bayesian analysis: <https://cran.r-project.org/web/views/Bayesian.html>)
6. It is an annotated introduction to the tools that are useful. We want to do something similar for forecasting.
7. The current EFI Task View is a big brain dump of tools.
8. The Overview - is Jake's attempt to organize the first big chunk of resources on the EFI Task View doc.
9. After we get what tools are available with the EFI Task View, the long term goal is to brainstorm where the gaps are and what tools are needed (e.g., downscaling climate data - have standard script so people don't need to continually create this script; bigger gap = JAGS/Stan not reading in their output)
10. Vision for intro blog post - layout in broad terms what the actual tasks are. There were 19 that were originally brainstormed in the EFI2019 DC meeting
11. We want a clean looking webpage that gives details about the resources.
12. The goal for the format of the website is up for discussion. Bulleted? 3-4 sentences for each resource?
  - a. Could vary on the level of depth that exists in each area.
  - b. Could be here is a concept and here are 5 R packages and 3 python scripts that match to that
  - c. Or could be describing what individual packages are providing
  - d. Give by table that can be sorted? Sort by R, python, etc

- e. Version 1 - start with text that can be ready for RCN (~6 weeks away)
  - f. Then in long term put it into a searchable system (similar to NEON or ORNL DAAC)
  - g. If we get to something that isn't easily accessible without sorting and metadata will be too much
  - h. Being opinionated and nudging people towards a set of tools that everyone uses is an okay thing to do
  - i. Want to restrict people that are new to forecasting to ways that are well supported
  - j. Another way to organize on the doc that we have now. Rather than starting with a list of tools, figure out what we want and what are the existing tools that can or can't do that.
13. Discussion of Jake's Overview doc
- a. Think about what we want, here is what is available and how it does/doesn't meet those needs. Include a principles section where we cite people
  - b. Eventually think about adding information about the best tutorials
  - c. Don't want to write a R vs Python vs Julia, then saying one of them is best. Want to lay out the distinctions between them that are fair to each program
  - d. Carl: From a user standpoint/those who are just getting into the field don't want to overwhelm them with all the options. They should pick whatever language people in their lab are using or people they are working with.
    - i. Convey throughout that the easiest thing to use is the thing that people around you are using
  - e. Could also think about this from a new users perspective - where and when are they most likely to encounter these things.
  - f. Forecasters will encounter compiled language when they are getting into mechanistic models or earth system models. If you are working with population models or aren't individual based will encounter scripted
  - g. Think why are we writing this. Giving beginners a tool to give advice? Or give sense of the forecasting universe? Two different things and make statements to guide readers. It is useful to have some acknowledgement that there are additional resources available. These are the resources out there that you need to be aware of. Do make it be a binary choice between programs.
  - h. Let new forecasters know that there are different languages that will interact with different tools.
  - i. We want to be able to reference things later. Want people to be aware of what Julia or compiled language means so introduce those now, so they are aware of them later.
  - j. Alexey to go through and strip out value judgements
  - k. Would be good to make sure to have the different metadata standards out there. Check with Hassan from NOAA has talked about a standard that NOAA has used that is different from XML
  - l. Carl tagged to help with the Metadata section. But could to have others join in to provide other disciplinary perspectives.

- m. Is there anything for people using Matlab? There is concern with Matlab because it is proprietary on the reproducibility. Be good to be open and acknowledge this at the beginning of the doc. We are filtering by tools that are open. Could also say there are packages in R, python, and Julia that can read Matlab objects. If you encounter Matlab script here are some things you can do with it even without running the proprietary. But put a big exclamation on it that says it is using proprietary software. Don't want to tell the community that you can't use a certain program. But can say "this document provides open source resources". Things we left out we are not putting a value judgement on, we just can't cover 100% of things.
  - n. Be good to to put in the program that reads Matlab code
14. Have we missed true workflow/automation tools?
- a. Don't see a traditional scientific workflow tools
  - b. Or made a workflow and want it to run in a scheduled way or a data driven way. Where do those tools fit into this structure? They are a key part of workflow for forecasting and it is one of the challenging leaps for the community - reproducible code, dealing with dependencies in docker.
    1. Split Continuous Integration and Code Testing
    2. Rename continuous integration to automation.
    3. Then have 2 types of tools. Those that run on other people's computers. And those that run on your own computer
    4. Carl likes the split
    5. This would be a good place for CS people to contribute. Rob and Kenton will weigh in on this section.
    6. Rob could give a list of 200 workflows. Pegasus (LIGO uses) and Airflow (because of Airbnb) are the two main workflows used most
    7. Apache open whisk is what Dietze's lab is using
    8. A Lot of this also overlaps with the containers in the containerization section

#### Data and Code Release

1. What are we envisioning for this?
2. Figure out the needs. We have talked about this. We just need to add here.
3. Needs to be searchable, updatable, size restrictions, archived stable.
4. Jody will go through the previous notes where we talked about this.

#### Moving forward

- Jake to continue to lead efforts to get this first doc where we can push it out as a blog post
- Try to do the draft iteration and send it out for everyone to look out and comment on before the next call. Want to get it out before/by the next Methods/CI call on April 17.
- Then start signing up folks to clean up the other sections in the other chunks of resources on the EFI Task View doc

- A note from the RCN steering committee call from earlier today. We want each working group to present what they have been doing at a high level. 10-15 minute lightning talk summarizing what each group has been doing. Want this on everyone's radar. Present what EFI has been doing over the past year before breakout groups.