December 9, 2022 Joint Methods & CI Working Group Call

Attendees: Matt Brousil, Libby Mohr, Carl Boettiger, Jody Peters
Regrets: Mike Dietze

Agenda/Notes:

1. Announcements
    a. Poll for scheduling Jan-May monthly calls.

    b. RCN June 21-23, 2023 Unconference at NEON HQ Save the Date
        i. Bringing People Together to Do Forecasting: Training, Technology, Theory, and Translation
        ii. The goal of the Unconference is for participants to work together to produce a product such as getting a forecast up and running, developing teaching materials, finalizing tutorials, refining or creating tools, analyzing forecasts for a manuscript, and/or developing visualizations. The event will also include a poster session for attendees to present their research.
        iii. Space is limited to 50 people. Short application will be available Jan 1 and due Feb 1. Applicants will be notified Feb 15.
        iv. Travel funds will be available for >30, <50

2. Emailed suggestion for Visualization Task View from Janis von Bleichert from experte.com
    a. I just saw that you refer to the accessibility checker wave.webaim.org on your site https://projects.ecoforecast.org/taskviews/visualization.html. I also think the tool is super useful, but unfortunately, only one single page can be tested at a time. So it can be very tedious to check your entire website for accessibility issues.
    b. That's why I have developed an accessibility checker that allows you to test hundreds of URLs automatically. After entering a URL, the tool crawls the website and determines the accessibility for each subpage in 8 categories. Some of the categories are e.g. correct labeling, contrast or ARIA attributes.
    c. The tool is free and can be used without registration at https://www.experte.com/accessibility.
    d. It would be great if you try the tool and, if you like it, add it as an additional resource to your page.

3. What are the interests of the group? What activity to focus on next? How to find a champion for specific activities?
    a. Mike: I can't make the call, but one thing to think about is how to leverage the enormous popularity of the EFI/ESA joint Methods webinar series when thinking about future activities. Are there ways to work on activities/events/products of interest to this broader audience? Are there activities that would bring that larger audience into a more involved role in the Methods/CI WG?

b. Matt can't help lead the Data Ingest Task View at this time as he onboards in his new role. Happy to be involved but can't lead it.
    i. Learning is an interest. Wanting to learn forecast skills. Discussions, journal club, workshop
c. Libby - joined the working group to learn what people are talking about in terms of methods and CI. Happy to contribute to the Task View
d. Carl - likes learning about new tools and democratizing tools so they are easy for people to run with - so they are accessible and usable across groups. Don't like it when it is siloed.
    i. Example of what Carl is working on.
        1. Forecasts are temporal. But think we need a Challenge that includes spatial analyses, so has been working with new tools and code for spatial analyses
        2. Cloud platform and cloud based workflows have changed how people work with spatial data
        3. Example: LAI from MODIS
            a. Used to have to work hard to extract data, now it is cloud based
            b. LAI from MODIS is available from STAC: https://stacspec.org/en.
            c. Carl gave example with the R code to show how to access LAI data from STAC
            d. Used library(rstac) and library(gdalcubes).
            e. Select the bounding box to get all the spatial LAI MODIS data from a specific area
            f. Assets are the images - they are tif images (used to be HDM5)
            g. There are 355 images
            h. Can construct a collection of images - it is fast because it doesn't download all the data. It just know it is available on the internet
            i. Data is sampled at 500 m resolution, but want to have it at the 0.1 of a degree (could go at 0.5 degree because that is the resolution of the NOAA GEFS weather forecast)
                i. Can specify the view independent of data source so can get resolution that matches across data sources
            j. Cube is in space and time and can make a gif animation
            k. If wanted to create a forecast and add a random walk to the layer then can add a pixel
            l. Created a gif for LAI over the year to see the change in greeness
        4. GEFS example - download grib files. Then can use the cube view to create an animation of the temp data for the first 10 days

    a. Carl has put together details for this example. It is a help package for the Forecast Challenge. You can see it here
        i. https://projects.ecoforecast.org/gefs4cast/articles/gefs-cog.html
5. Is it all running in the cloud?
    a. Yes and No.
    b. Old school - the CPU had to be in the cloud. Now the CPU is on Carl's desk. Instead of CPU talking to your harddrive your CPU is going over the network. It will all run with almost no RAM because it is put on the network which is very fast. Cloud provider just needs to give storage that is plugged into fast network.
    c. Compare this to the way Googles earth Engine works - yo send your code off to run on their servers and access their data and sends its results back to you.
    d. Here the data goes back and forth over the network, but not tied to their API
6. Have done the same thing with the parquet - data sits in the static buckets
        i. In R - you use read csv and it reads the data in and if it is too big it crashes
        ii. Could use dplyr with data on your local disk instead to read in a the data
        iii. parquet is doing the same sort of thing
        iv. Instead of csv it is binary and automatically compresses.
        v. Don't have to worry about dates becoming integers or characters
        vi. Can read in just the part of the data you want. You don't have to get the full set of data
        vii. How does it know what range to request without the data?
            1. When you read it into memory it is converted into a format it understands. So the trick is to have software that can do it without running it on RAM
            2. It is faster for the computer to read it in columns instead of rows
        viii. Is it like Apache Arrow but for geospatial data?
            1. Yes
7. Example with STARS
    a. Using Leaf Area asset. Grab the URL
    b. Can use the remote network instead of the local disk

        c. Complaint that students don't have paths on their computer to read the csvs. This addresses this so students don't need to worry about setting up the file path on their computers.
        d. This lets you get a file that is on the internet. It doesn't have the same metadata with last time modified and the permissions. So it is a new paradigm
    ii. What is the data cube?
        1. Data cube is x, y and time . This is also called a mosaic and can tile them in space and deal with the overlap
    iii. On demand cubes is new. Before had to precompute the cubes and others would download and do the analysis you did. But because of different resolutions and needs, wondering how much to provide.
    iv. Need to do this more - have more people talk together. Carl wants to be the bridge. If people have questions, post them in Slack and Carl will help out (and post on other Slacks he is on :-))

4. Previous Project Ideas
    a. Data Ingest, Cleaning, and Management Task View
    b. Forecasting Standards and Challenge CI Update
    c. Workshop Proposal
        i. Google doc to collectively develop ideas
    d. Forecast visualization pitch for the NEON Challenge (this came up on the Sept call, think about if this is something the group wants to move forward with)