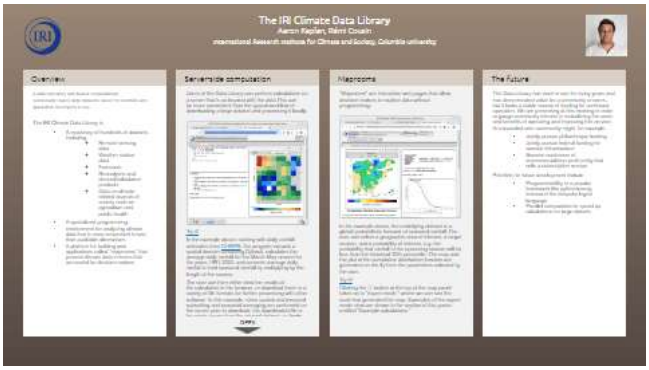
The IRI Climate Data Library



Aaron Kaplan, Rémi Cousin

International Research Institute for Climate and Society, Columbia University

PRESENTED AT:

OVERVIEW

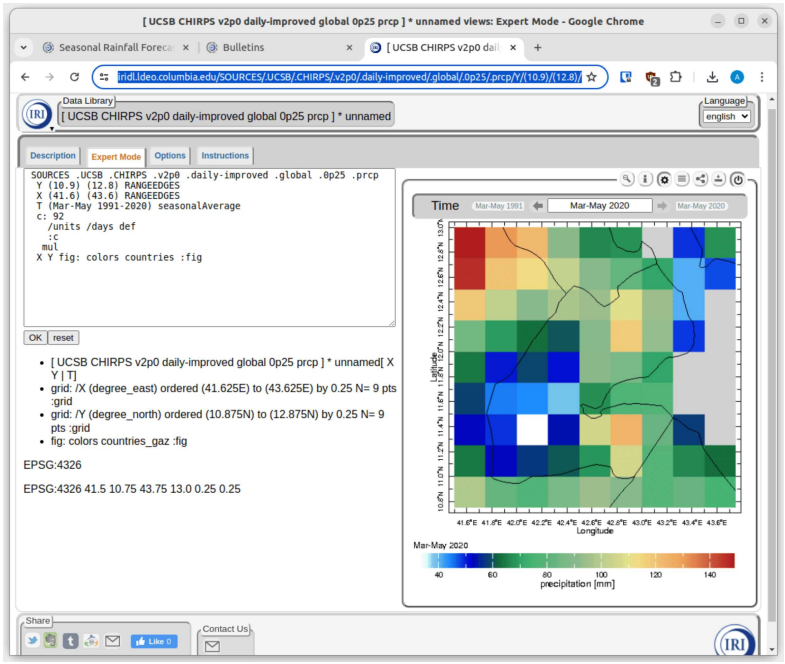
A data repository with built-in computational functionality makes large datasets easier for scientists and application developers to use.

The IRI Climate Data Library is

- A repository of hundreds of datasets including
  - Remote sensing data
  - Weather station data
  - Forecasts
  - Reanalyses and derived/calculated products
  - Data on climate-related aspects of society such as agriculture and public health.
- A specialized programming environment for analyzing climate data that is more convenient to use than available alternatives
- A platform for building web applications called "maprooms" that present climate data in forms that are useful for decision makers

SERVER-SIDE COMPUTATION

Users of the Data Library can perform calculations on a server that is co-located with the data.This can be more convenient than the typical workflow of downloading a large dataset and processing it locally.



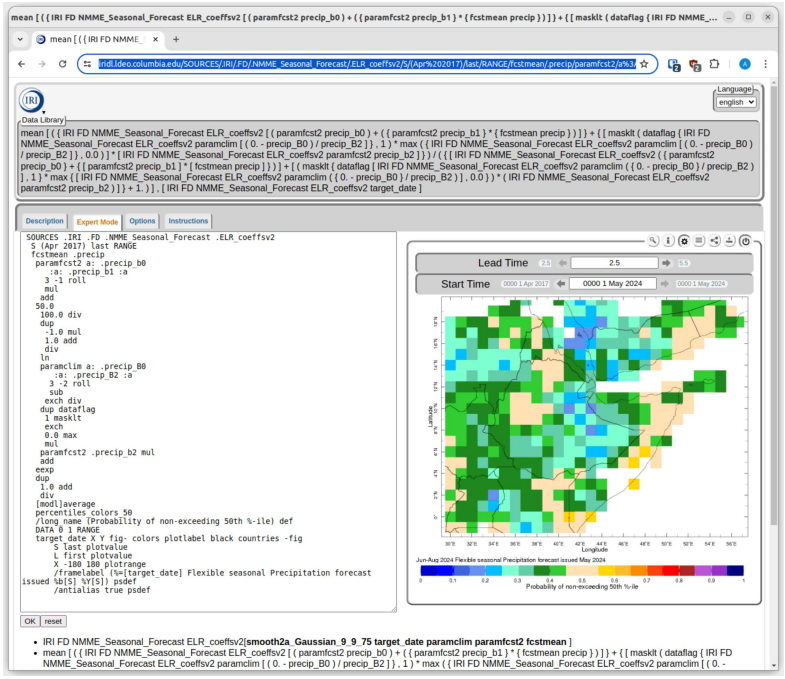
Try it! (http://iridl.ldeo.columbia.edu/SOURCES/.UCSB/.CHIRPS/.v2p0/.daily-improved/.global/.0p25/.prcp/Y/(10.9)/(12.8)/RANGEEDGES/X/(41.6)/(43.6)/RANGEEDGES/T/(Mar-May%201991-2020)/seasonalAverage/c%3A/92//units//days/def/%3Ac/mul/X/Y/fig%3A/colors/countries/%3Afig/?T=Mar-May%202020#expert)

In the example above, starting with daily rainfall estimates from CHIRPS (https://www.chc.ucsb.edu/data/chirps), the program extracts a spatial domain containing Djibouti, calculates the average daily rainfall for the March-May season for the years 1991-2020, and converts average daily rainfal to total seasonal rainfall by multiplying by the length of the season.

The user can then either view the results of the calculation in the browser, or download them in a variety of file formats for further processing with other software. In this example, since spatial and temporal subsetting and seasonal averaging are performed on the server prior to download, the downloaded file is far easier to use than the original dataset: **a single 9KB file, compared to 33GB spread over 15,000 files.**

The programming instructions for a calculation or visualization are encoded as a URL, which makes it easy to share code and results with others, or to fetch data programatically.

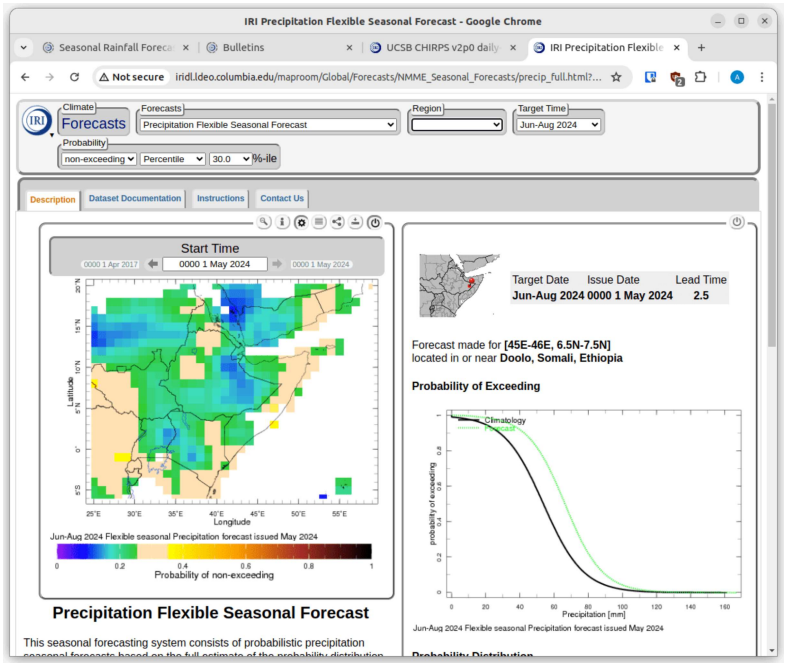
The following program plots the probability of exceedance of the 50th historical percentile according to a probabilistic forecast generated using extended logistic regression (ELR [1]). The input to the calculation is the forecast expressed as a set of ELR coefficients for each combination of (latitude, longitude, initialization date, lead time).



Try it!
(http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD/.NMME_Seasonal_Forecast/.ELR_coeffsv2/S/(Apr%202017)/last/RANGE/fcstmean/.precip/paramfcst2/a%3A/.precip_b0/%
exceeding%2050th%20%25-ile)/def/DATA/0/1/RANGE/target_date/X/Y/fig-/colors/plotlabel/black/countries/-
fig/S/last/plotvalue/L/first/plotvalue/X/-180/180/plotrange//framelabel/(%25%3D%5Btarget_date%5D%20Flexible%20seasonal%20Precipitation%20forecast%20issued%20%2
bbox=bb%3A29.52%3A-1.50%3A57.17%3A19.99%3Abb#expert)

MAPROOMS

"Maprooms" are interactive web pages that allow decision makers to explore data without programming.



In the example above, the underlying dataset is a global probabilistic forecast of seasonal rainfall. The user can select a geographic area of interest, a target season, and a probability of interest, e.g. the probability that rainfall in the upcoming season will be less than the historical 30th percentile. The map and the plot of the cumulative distribution function are generated on the fly from the parameters selected by the user.

Try it! (http://iridl.ldeo.columbia.edu/maproom/Global/Forecasts/NMME_Seasonal_Forecasts/precip_full.html)

Clicking the ⓘ button at the top of the map panel takes us to "expert mode," where we can see the code that generated the map. Examples of the expert mode view are shown in the section of this poster entitled "Example calculations."

THE FUTURE

The Data Library has been in use for many years and has demonstrated value for a community of users, but it lacks a stable source of funding for continued operation. We are presenting at this meeting in order to gauge community interest in mutualizing the costs and benefits of operating and improving this service. An expanded user community might, for example,

- Jointly pursue philanthropic funding
- Jointly pursue federal funding for science infrastructure
- Become customers of a commercial/non-profit entity that sells a subscription service

Priorities for future development include

- Programmability in a popular framework like python/xarray, instead of the bespoke Ingrid language
- Parallel computation to speed up calculations on large datsets

AUTHOR INFORMATION

Aaron Kaplan, Rémi Cousin

International Research Institute for Climate and Society

Climate School, Columbia University

help@iri.columbia.edu

TRANSCRIPT

REFERENCES

1. Extending logistic regression to provide full-probability-distribution MOS forecasts. Daniel S. Wilks. Meteorol. Appl. 16: 361–368 (2009). DOI: 10.1002/met.134