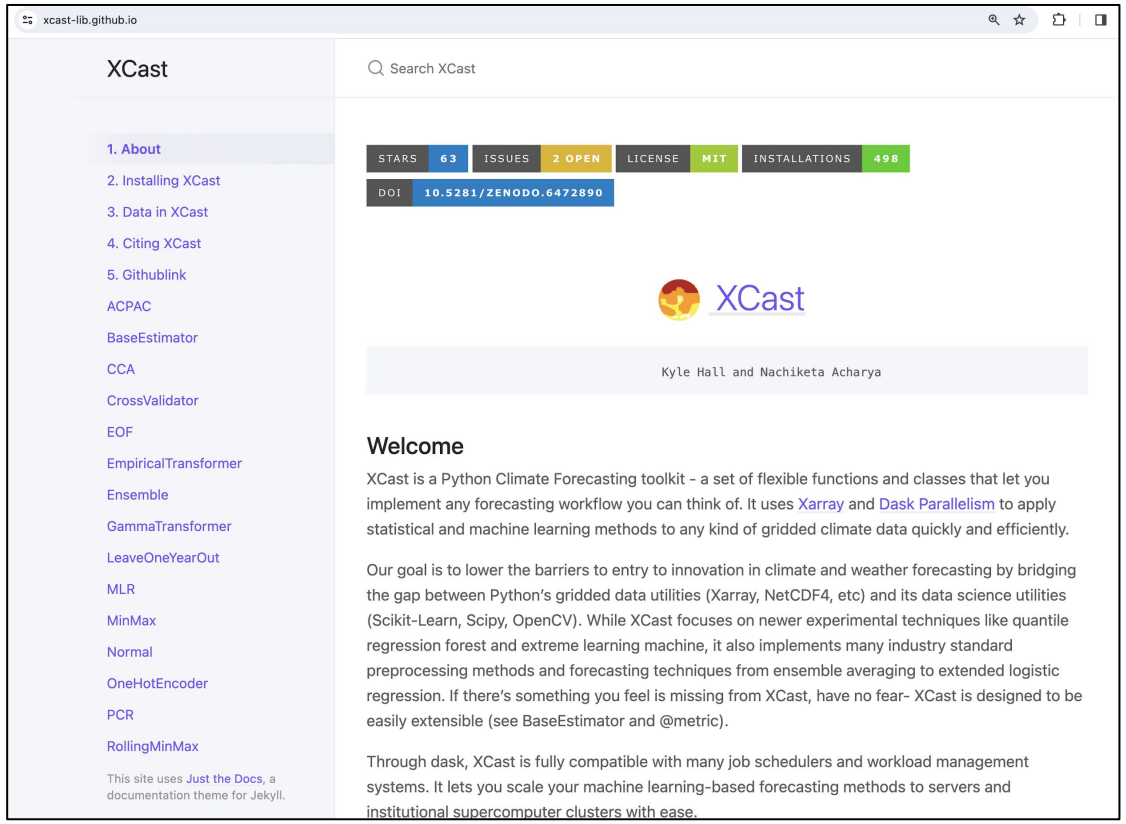
XCast: An AI/ML based S2S forecasting tool for Climate Services

Nachiketa Acharya

Cooperative Institute for Research in the Environmental Sciences (CIRES), University of Colorado Boulder

NOAA Physical Sciences Laboratory, Boulder, Colorado, USA

CIRES
University of Colorado Boulder



https://xcast-lib.github.io/

Collaborator: Kyle Hall

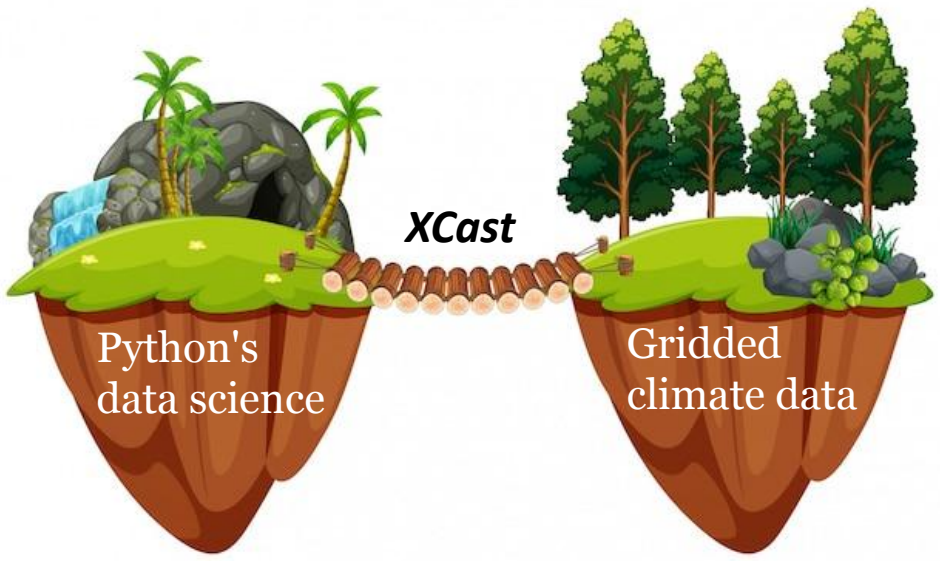
Why we need another tool?

Using AI/ML calibration techniques on S2S-scale GCM outputs may be exciting, but it also presents a number of barriers to entry, such as:

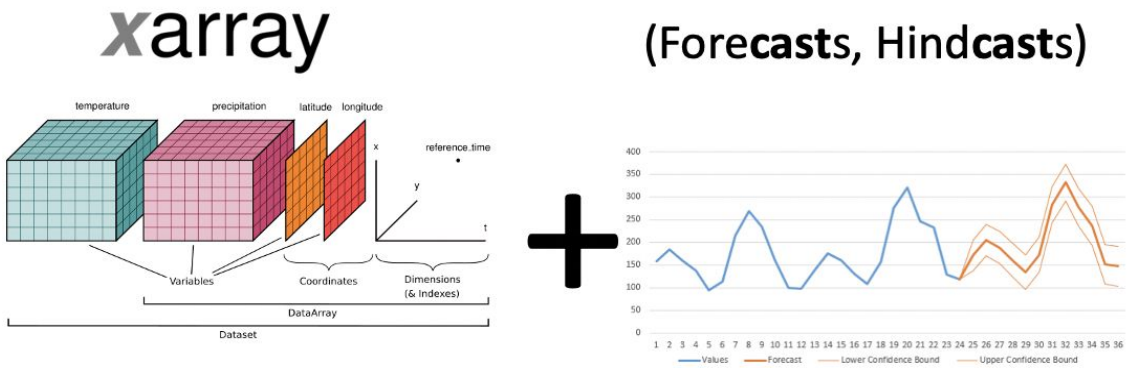
- Complex Data Pre-Processing: The diversity of configurations and formats across S2S forecast databases necessitates a high level of flexibility which no tool has yet achieved.
- Incompatibility of Python ML Packages with Geospatial Data: Extracting data from climate data formats (NetCDF, Grib2) in order to use it with Python ML packages presents an additional preprocessing step.
- Slow Algorithms: The complexity of many AI/ ML algorithms presents a problem since model training time and resources required can scale nonlinearly with data volume.
- Limitations Of Traditional Tools: Traditional climate forecasting tools implement methods relying on Gaussian processes, but week 3-4 precipitation forecasts and other S2S target variables do not follow Gaussian distributior

What is *XCast* ?

XCast, is a Python based High-Performance Data Science Toolkit for Climate Forecasting



Bridge the gap between Python's data science ecosystem and gridded climate data ecosystem



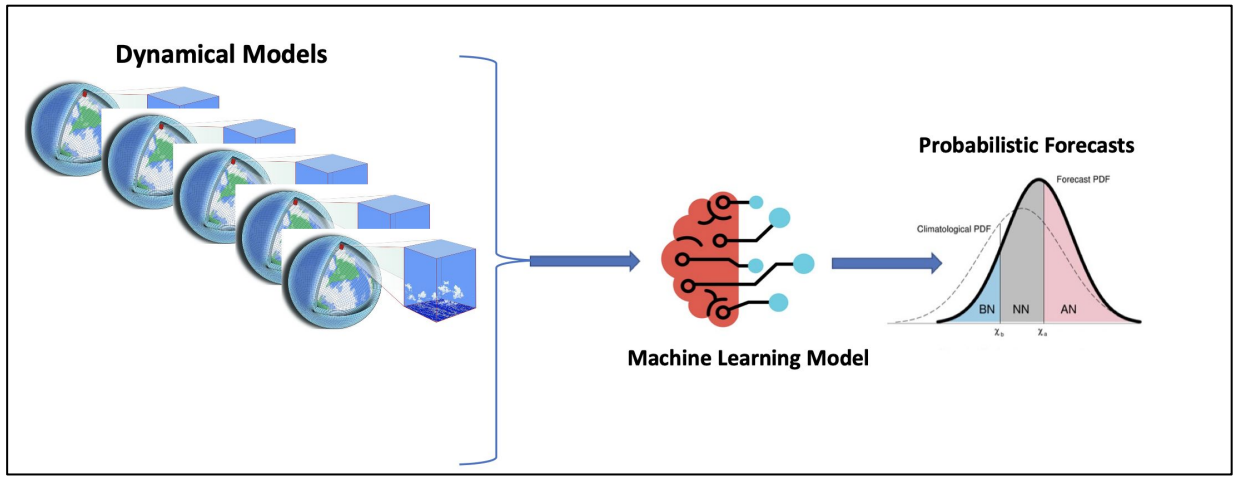
Core component:

- Xarray to handle multi-dimensional data.
- Estimators (Regression, Classification).
- Validation (Leave-N-Out Cross Validator).
- Skill Metrics (Correlation, Skill Scores).
- Pre and Post-Processing.
- Parallel processing using *Dask*.

Advantage *XCast* tool over traditional tools(CPT,PyCPT etc.)

- Lots of Python package availed but no one can handle multi-dimensional for climate dataset.
- It is super easy to installed (one line command), no need any special computer and it works for Window, Mac and Linux/Unix system.
- Xcast parallelize the code, so it is much faster.
- It includes all the traditional methods in (MLR,PCR,CCA) and most advanced AI/ML methods (like ANN, Random Forest etc.)
- It read NetCDF/Grib2/Zar data where traditional tool needs “ASCII format”.
- It can also read any model outputs (NMME,C3S,S2S and SubX or your own).
- It is not just a “Jupyter notebook” rather it is a Python Package.

History of *XCast* Development



TITLE	CITED BY	YEAR
Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme …	121	2014
N Acharya, NA Shrivastava, BK Panigrahi, UC Mohanty		
Climate dynamics 43, 1303-1310		

Clim Dyn (2014) 43:1303–1310
DOI 10.1007/s00382-013-1942-2

Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine

Nachiketa Acharya · Nitin Anand Shrivastava · B. K. Panigrahi · U. C. Mohanty

Received: 9 May 2013 / Accepted: 10 September 2013 / Published online: 20 September 2013

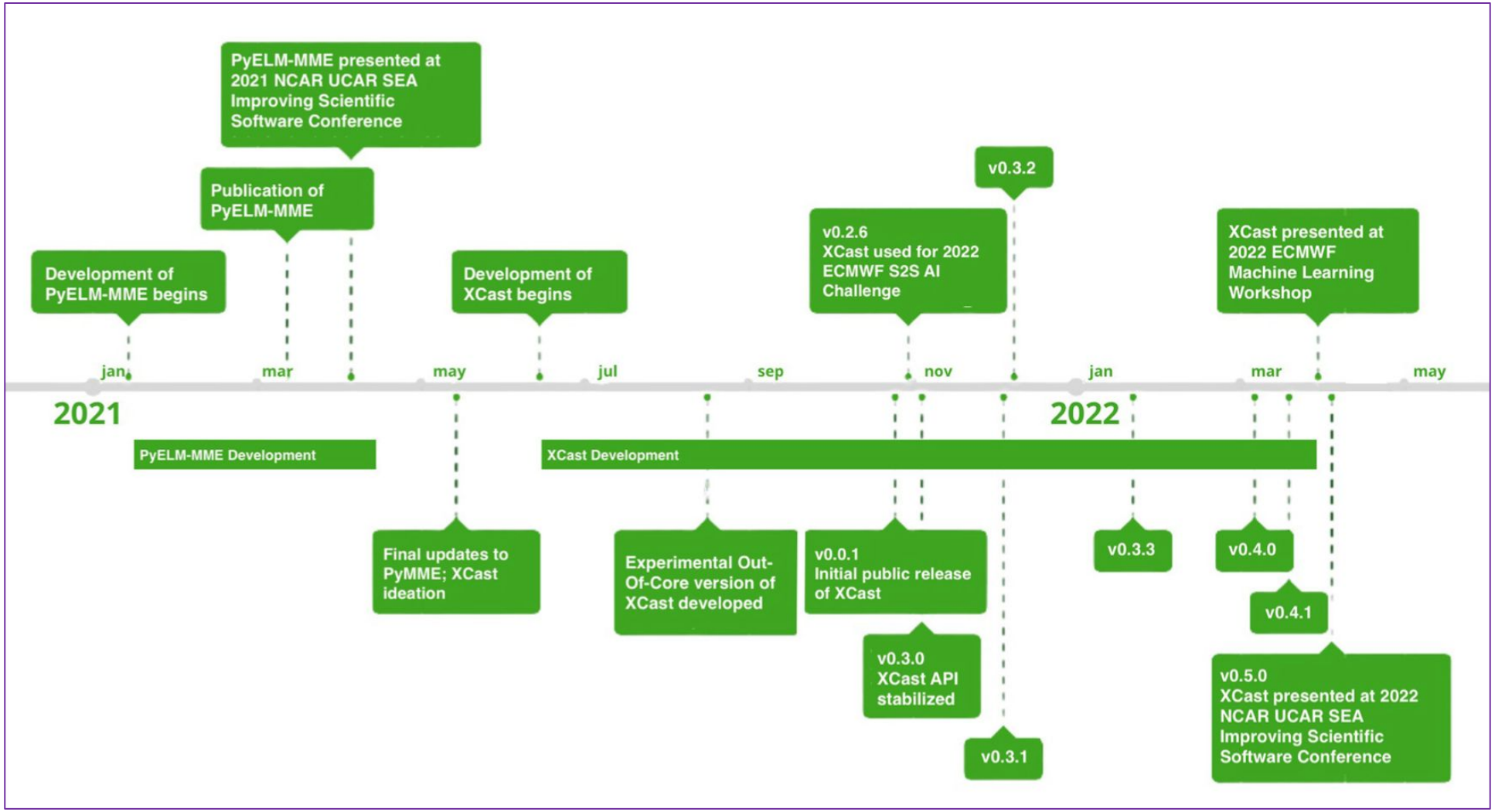
Abstract The south peninsular part of India gets maximum amount of rainfall during the northeast monsoon (NEM) season [October to November (OND)] which is the primary source of water for the agricultural activities in this region. A nonlinear method viz., Extreme learning machine (ELM) has been employed on general circulation model (GCM) products to make the multi-model ensemble (MME) based estimation of NEM rainfall (NEMR). The ELM is basically is an improved learning algorithm for the single feed-forward neural network (SLFN) architecture. The 27 year (1982–2008) lead-1 (using initial conditions of September for forecasting the mean rainfall of OND) hindcast runs (1982–2008) from seven GCM has been used to make MME. The improvement of the proposed method with respect to other regular MME (simple arithmetic mean of GCMs (EM) and singular value decomposition based multiple linear regressions based MME) has been assessed through several skill metrics like Spread distribution, multiplicative bias, prediction errors, the yield of prediction, Pearson's and Kendal's correlation coefficient and Wilmort's index of agreement. The efficiency of ELM estimated rainfall is established by all the stated skill scores. The performance of ELM in extreme NEMR years, out of which 4 years are characterized by deficit rainfall and 5 years are identified as excess, is also examined. It is found that the ELM could expeditiously capture these extremes reasonably well as compared to the other MME approaches.

Keywords Northeast monsoon rainfall · General circulation models · Multi-model ensemble · Artificial neural network · Extreme learning machine

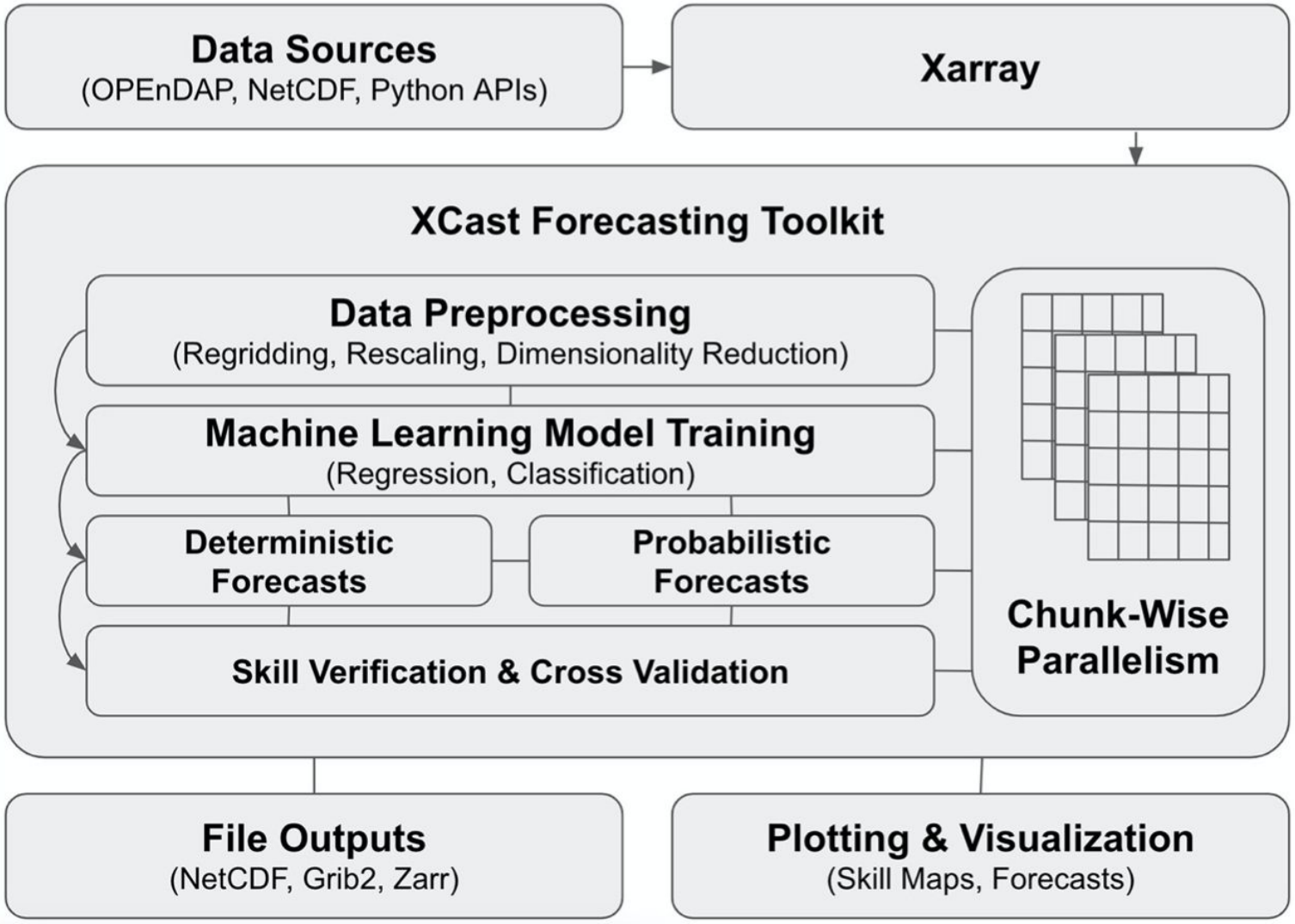
1 Introduction

Though the southwest monsoon (SWM) during the months of June to September (JJAS) is the principal rainfall season of India, the south peninsular part of the country does not get much rainfall as this region falls under the rain shadow region. The south peninsular region comprises the meteorological subdivisions of Coastal Andhra Pradesh, Rayalaseema and Tamilnadu-Pondicherry experience major rainfall activity during the period of October–November–December (OND) which is referred to as the northeast monsoon (NEM) or the post monsoon season. Many of the above subdivisions receive 17–49 % of their annual rain-

History of *XCast* Development



How *XCast* works?



How *XCast* works?

https://www.frontiersin.org/articles/10.3389/fclim.2022.953262/full

ORIGINAL RESEARCH article

Front. Clim., 15 July 2022
Sec. Predictions and Projections
Volume 4 - 2022 | https://doi.org/10.3389/fclim.2022.953262

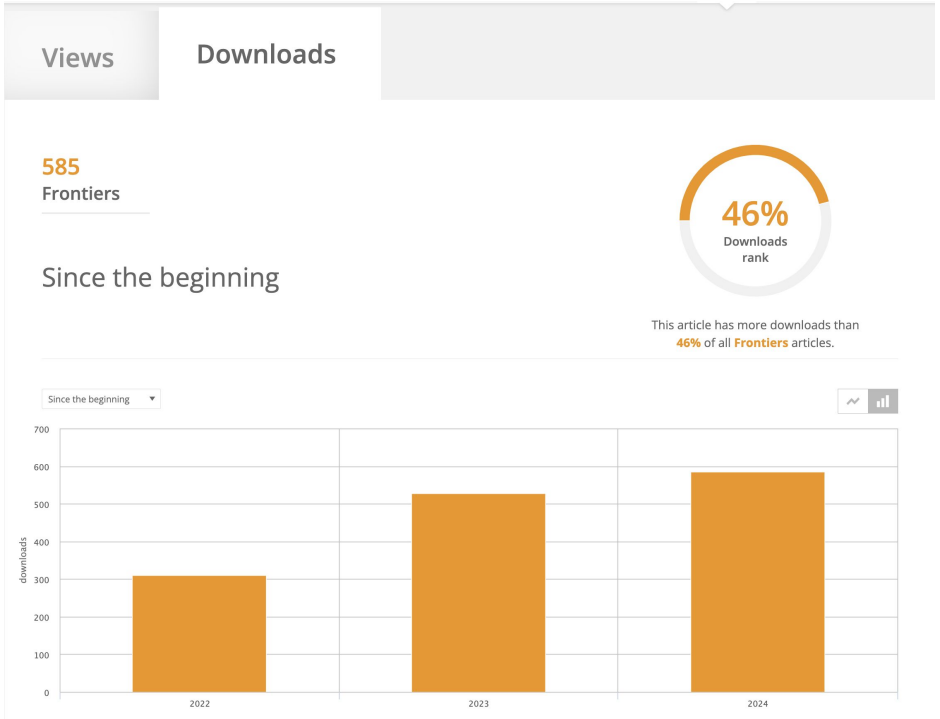
This article is part of the Research Topic
Machine Learning for Climate Predictions and Projections
View all 6 Articles ›

XCast: A python climate forecasting toolkit

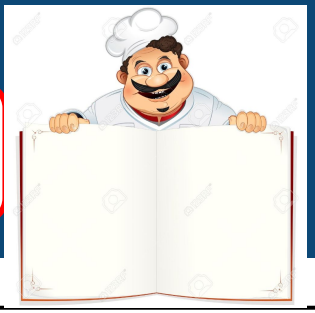
 Kyle Joseph Chen Hall[1*]  Nachiketa Acharya[2,3]

[1] International Research Institute for Climate and Society, Columbia University, Palisades, NY, United States
[2] Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, United States
[3] National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory, Boulder, CO, United States

Climate forecasts, both experimental and operational, are often made by calibrating Global Climate Model (GCM) outputs with observed climate variables using statistical and machine learning models. Often, machine learning techniques are applied to gridded data independently at each gridpoint. However, the implementation of these gridpoint-wise operations is a significant barrier to entry to climate data science. Unfortunately, there is a significant disconnect between the Python data science ecosystem and the gridded earth data ecosystem. Traditional Python data science tools are not designed to be used with gridded datasets, like those commonly used in climate forecasting. Heavy data preprocessing is needed: gridded data must be aggregated, reshaped, or reduced in dimensionality in order to fit the strict formatting requirements of Python's data science tools. Efficiently implementing this gridpoint-wise workflow is a time-consuming logistical burden which presents a high barrier to entry to earth data science. A set of high-performance, easy-to-use Python climate forecasting tools is needed to bridge the gap between Python's data science ecosystem and its gridded earth data ecosystem. XCast, an Xarray-based climate forecasting Python library developed by the authors, bridges this gap. XCast wraps underlying two-dimensional data science methods, like those of Scikit-Learn, with data structures that allow them to be applied to each gridpoint independently. XCast uses high-performance computing libraries to efficiently parallelize the gridpoint-wise application of data science utilities and make Python's traditional data science toolkits compatible with multidimensional gridded data. XCast also implements a diverse set of climate forecasting tools including traditional statistical methods, state-of-the-art machine learning approaches, preprocessing functionality (regridding, rescaling, smoothing), and postprocessing modules (cross validation, forecast verification, visualization). These tools are useful for producing and analyzing both experimental and operational climate forecasts. In this study, we describe the development of XCast, and present in-depth technical details on how XCast brings highly parallelized gridpoint-wise versions of traditional Python data science tools into Python's gridded earth data ecosystem. We also demonstrate a case study where XCast was used to generate experimental real-time deterministic and probabilistic forecasts for South Asian Summer Monsoon Rainfall in 2022 using different machine learning-based multi-model ensembles.



XCast's Menu Card



Pre-processor

- Interpolation
- Scaling
- Dimensionality Reduction
- One-Hot Encoding
- Many more

Deterministic Forecasts

- Bias Corrected Ensemble Mean
- Multiple Linear Regression
- Principal Component Regression
- Canonical Correlation Analysis
- Ridge Regression
- Extreme Learning Machine (ELM)
- Multi-Layer Perceptron
- Random Forest
- Many more

Probabilistic Forecasts

- Counting Member
- Prob.Anomaly Correlation(PAC)
- Extended Logistic Regression
- Multivariate Logistic Regression
- Probabilistic Output ELM
- **Extended Probabilistic Output ELM (EPO-ELM)**
- Naïve Bayes
- Quantile Random Forest
- Many more

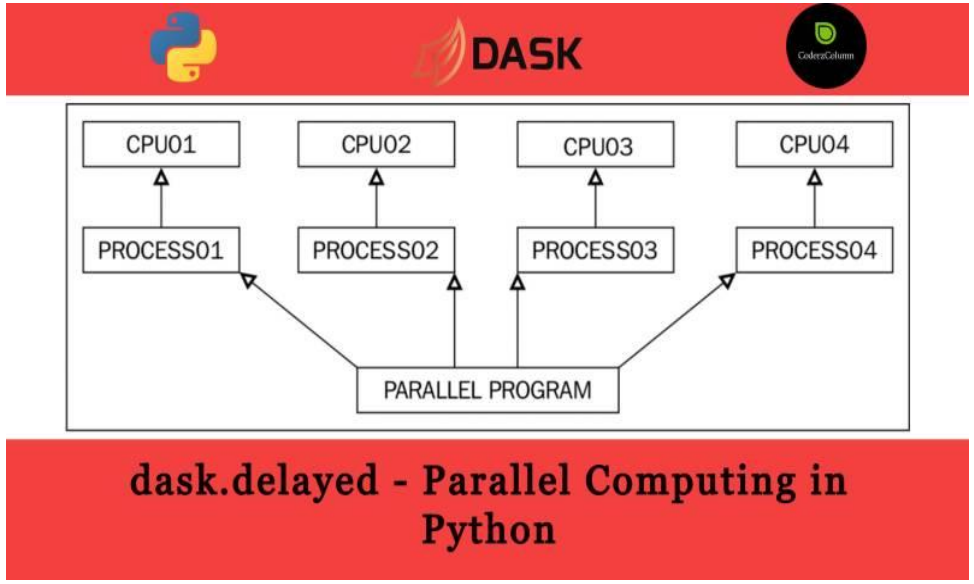
Verification Metrics

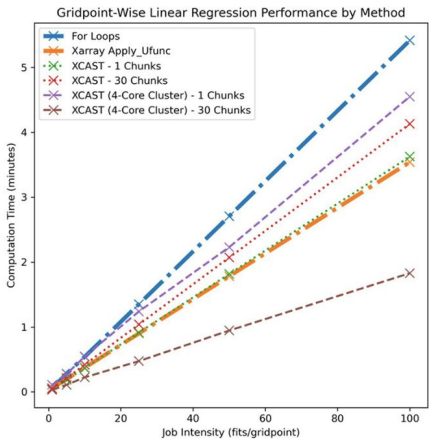
- Index of Agreement
- Nash-Sutcliffe Efficiency
- Kling-Gupta Efficiency
- Point-Biserial Correlation
- Brier Skill Score
- Continuous RPSS
- Generalized ROC
- Ignorance Score
- All traditional metrics

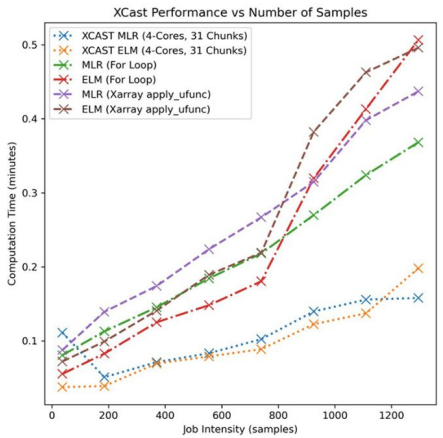
For more look at the Appendix C.1 in Hall and Acharya,2022

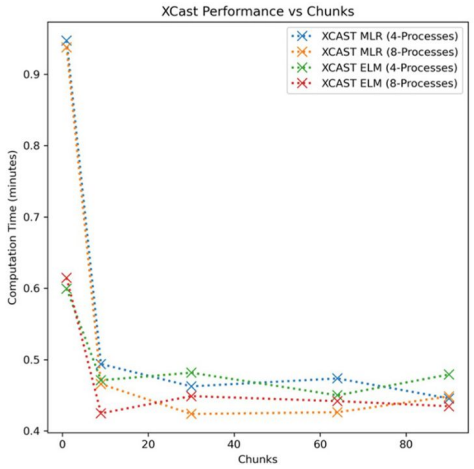
Parallelism

- *XCast* uses Dask Futures.









Parallelism: Example for ELR

```
(xcastv1) khall@khall-MBP endalk % python gen_elr-threadwise_inmemory.py
       START: 2024-01-08 12:00:06.094364
        OPEN: 0:00:00.267615 (elapsed)
     DRYMASK: 0:00:13.778541 (elapsed)
MISSING MASK: 0:00:00.450516 (elapsed)
      REGRID: 0:00:00.158656 (elapsed)
xval step 60
AVG XVAL STEP: 0:00:08.824152 (elapsed)
     ONE-HOT: 0:00:07.126371 (elapsed)
UNABLE TO FIND NAMES FOR FOLLOWING DIMS: ['sample', 'feature']
UNABLE TO FIND NAMES FOR FOLLOWING DIMS: ['sample', 'feature']

TOTAL ELAPSED TIME:  0:09:19.621245
```

Dask one process

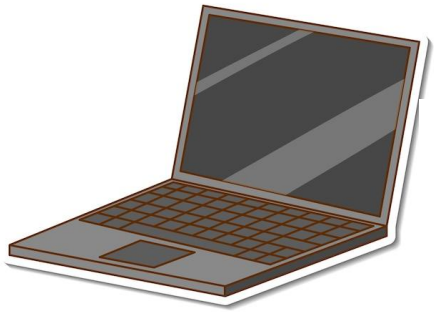
Dask multi-processing with XCast yields about a 3x speed up!

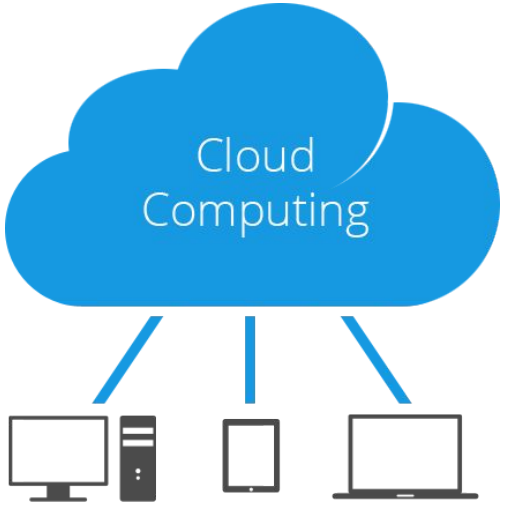
```
(xcastv1) khall@khall-MBP endalk % python gen_elr-dask_processwise.py
       START: 2024-01-08 12:09:48.971690
        OPEN: 0:00:00.396763 (elapsed)
     DRYMASK: 0:00:00.099413 (elapsed)
MISSING MASK: 0:00:00.930284 (elapsed)
      REGRID: 0:00:00.283437 (elapsed)
xval step 60
AVG XVAL STEP: 0:00:02.803414 (elapsed)
     ONE-HOT: 0:00:00.226898 (elapsed)
/Users/khall/miniconda3/envs/xcastv1/lib/python3.10/site-packages/distributed/client.py:3095: UserWarning: Sending large graph of size 33.50 MiB.
This may cause some slowdown.
Consider scattering data ahead of time and using futures.
  warnings.warn(
UNABLE TO FIND NAMES FOR FOLLOWING DIMS: ['sample', 'feature']
/Users/khall/miniconda3/envs/xcastv1/lib/python3.10/site-packages/distributed/client.py:3095: UserWarning: Sending large graph of size 33.50 MiB.
This may cause some slowdown.
Consider scattering data ahead of time and using futures.
  warnings.warn(
UNABLE TO FIND NAMES FOR FOLLOWING DIMS: ['sample', 'feature']
TOTAL ELAPSED TIME:  0:03:23.721158
```

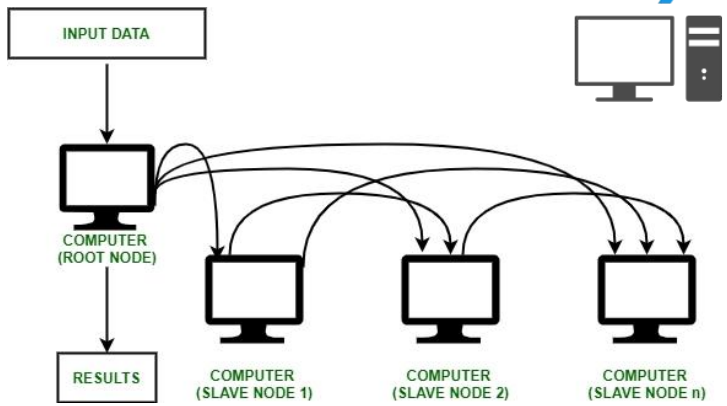
Dask multiprocessing

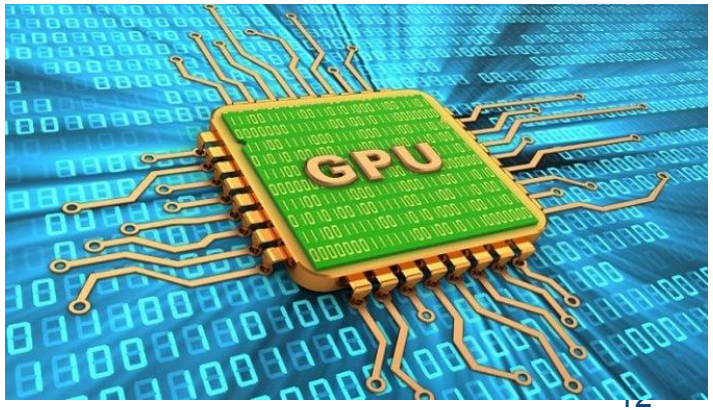
Running Platform

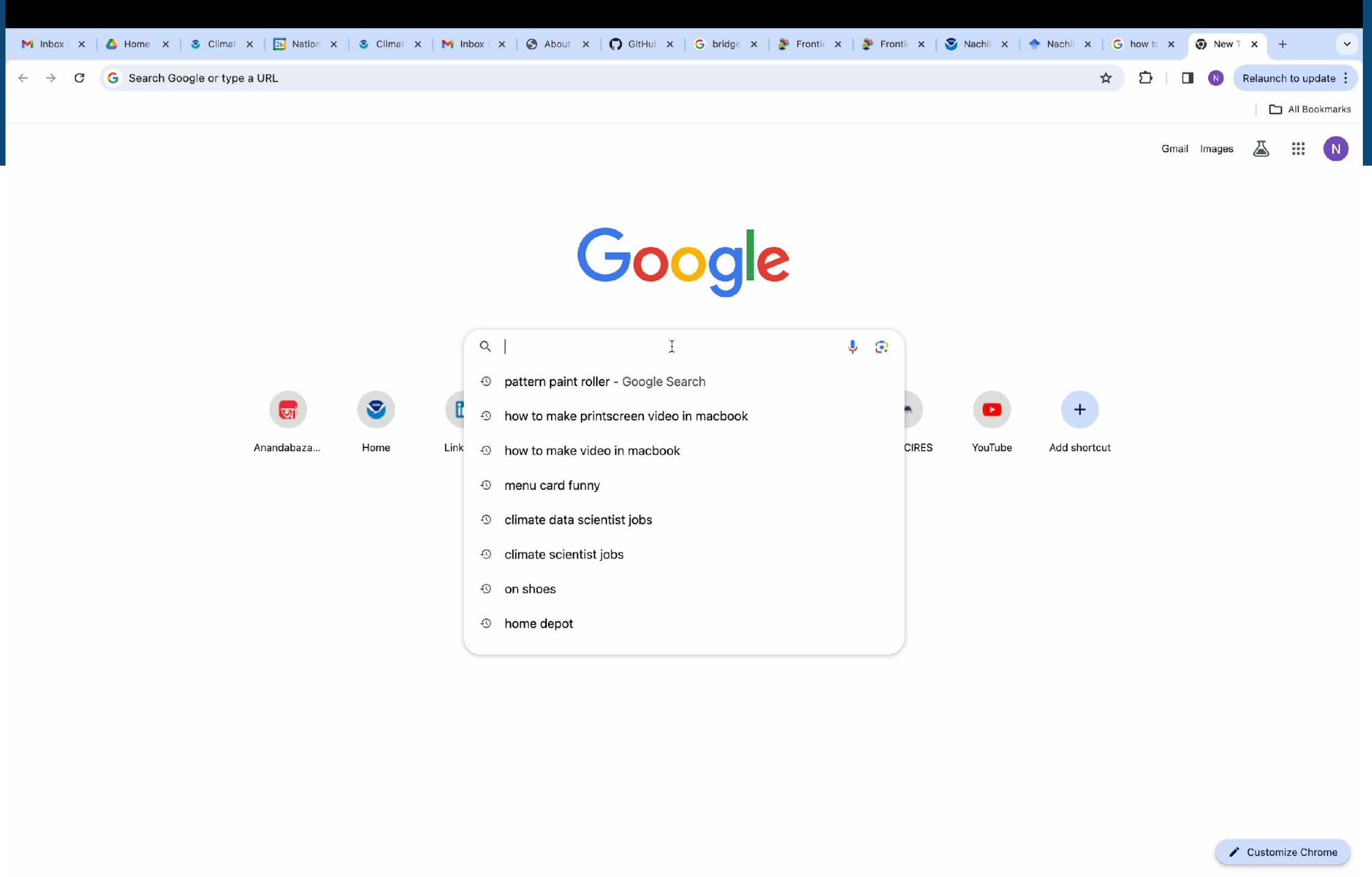
- *XCast* can run on Laptop, Work Station, Cluster, Cloud and GPU.



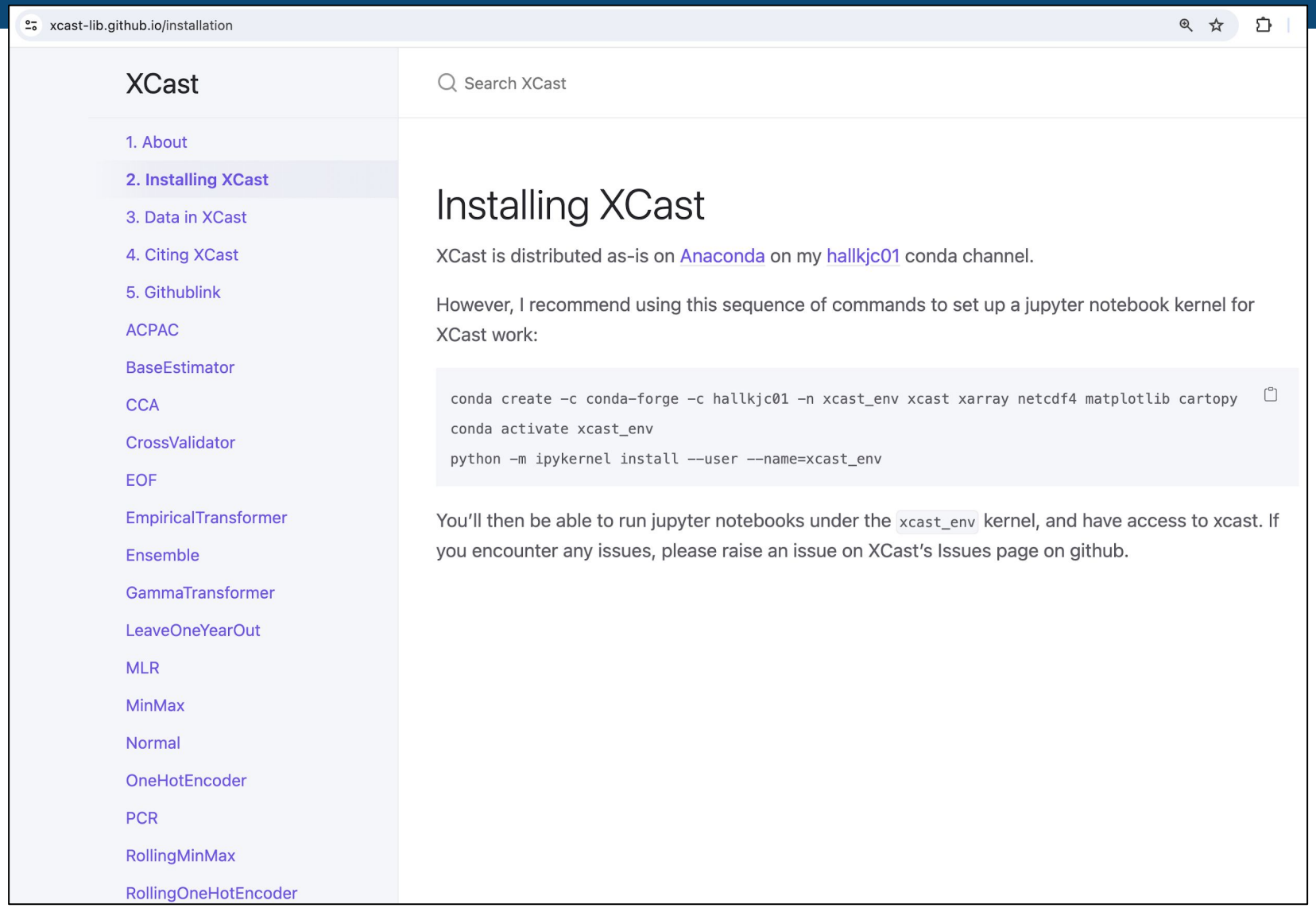




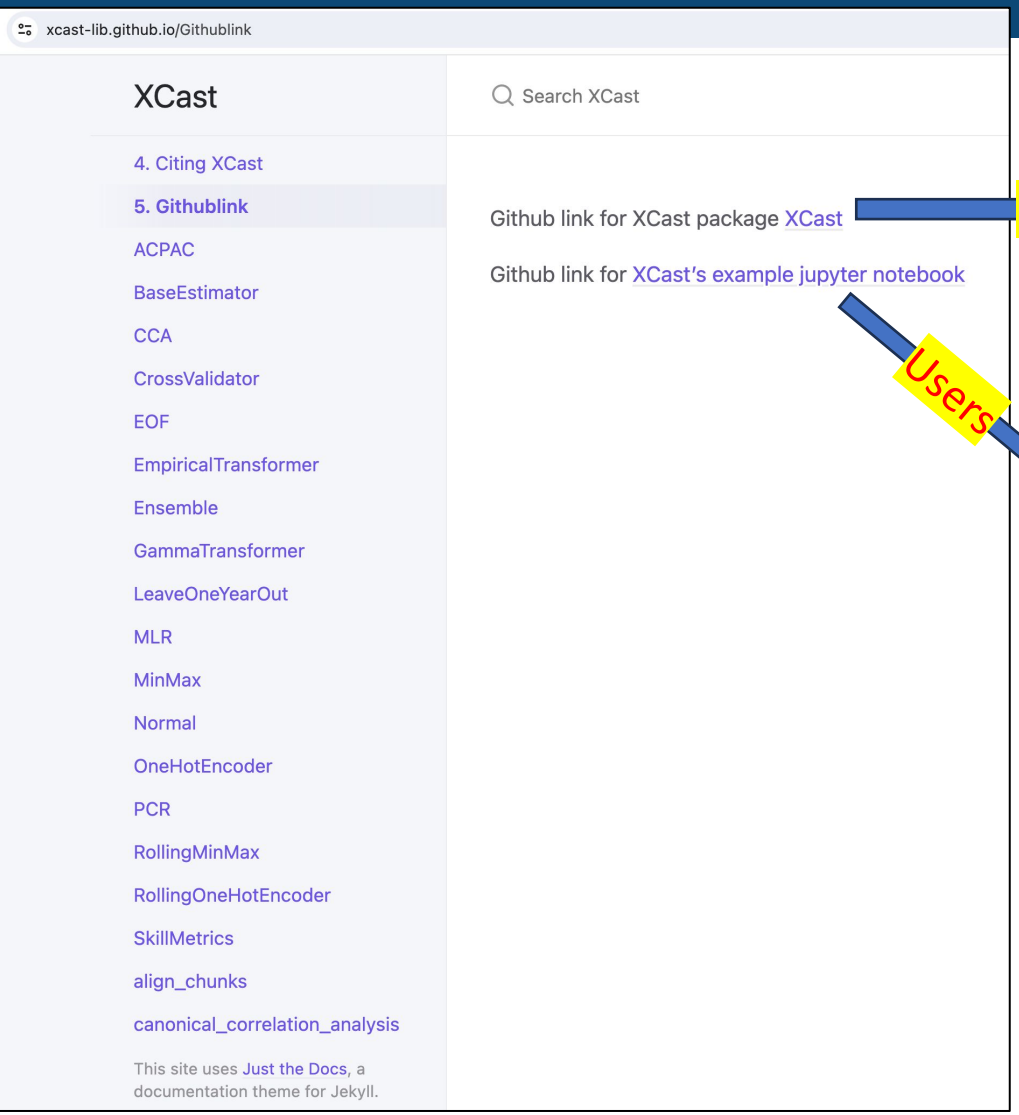


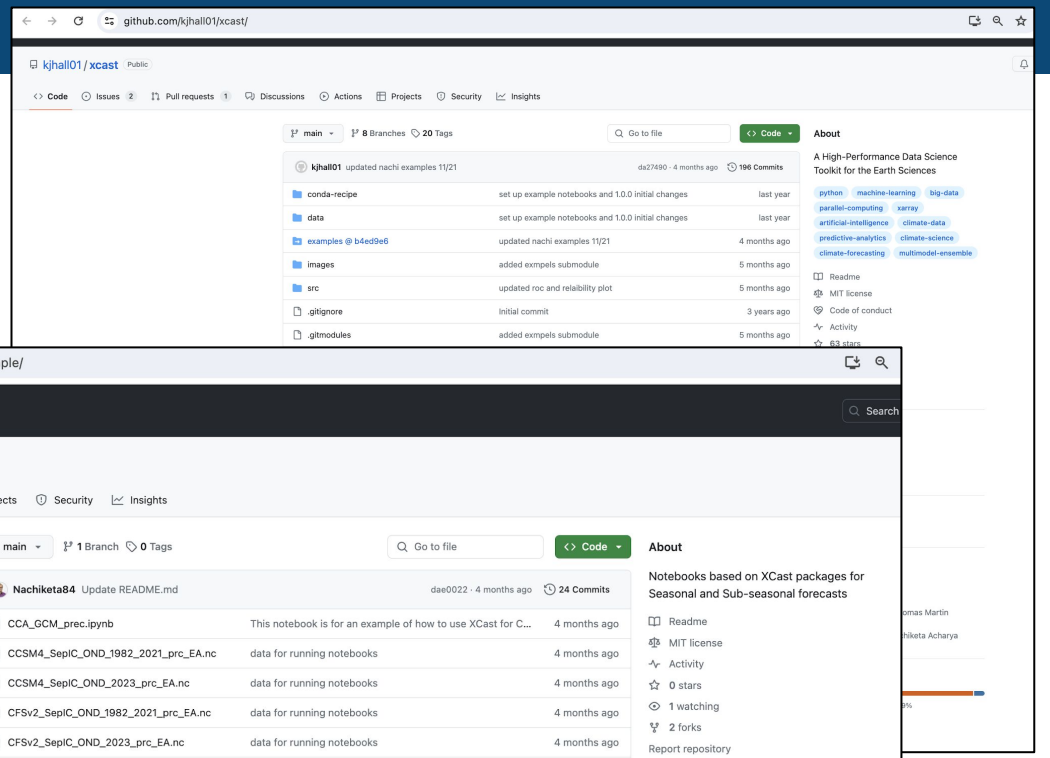

Search Google or type a URL
Relaunch to update
All Bookmarks
Gmail
Images
Google
pattern paint roller - Google Search
how to make printscreen video in macbook
how to make video in macbook
menu card funny
climate data scientist jobs
climate scientist jobs
on shoes
home depot
Anandabaza...
Home
YouTube
Add shortcut
Customize Chrome

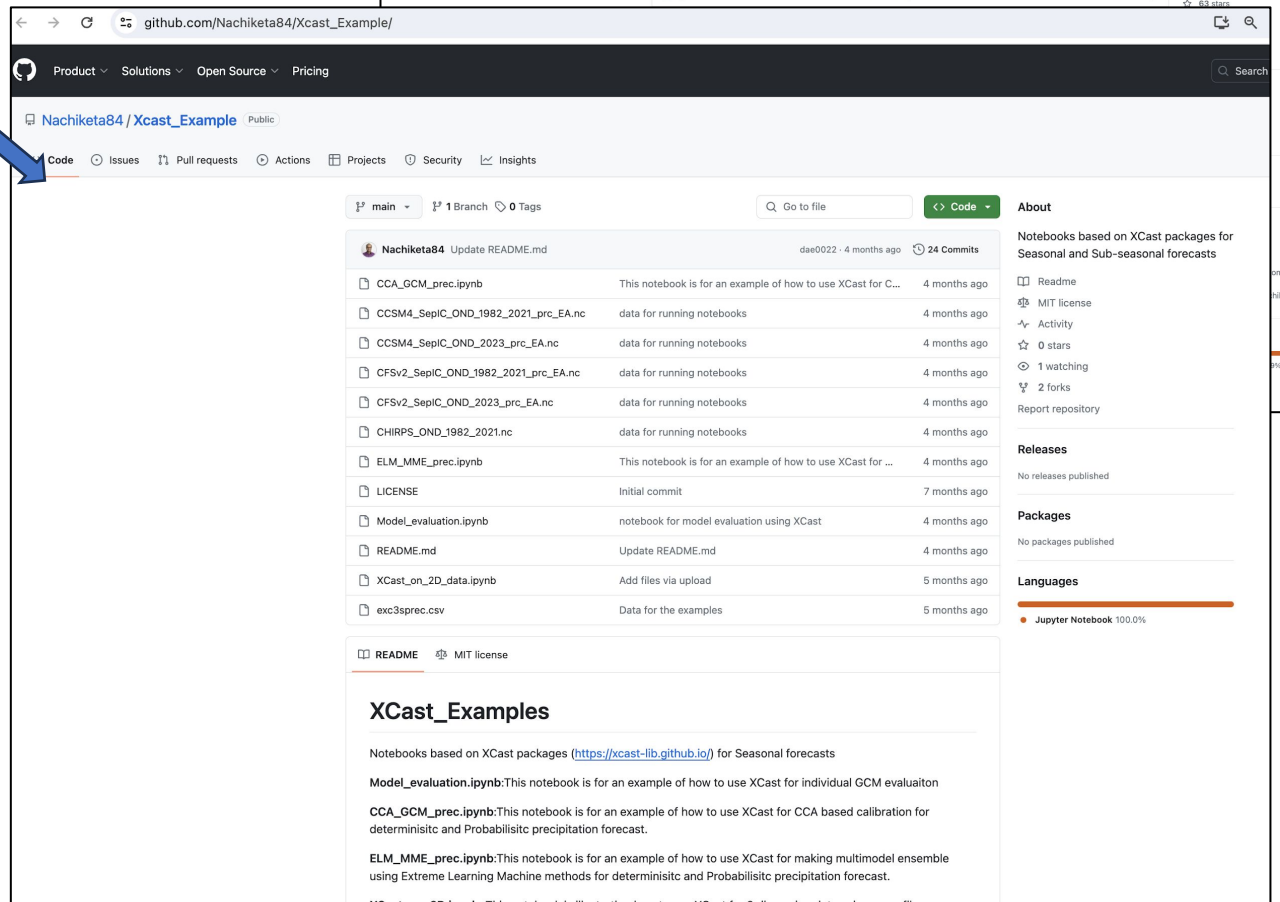
Installing *XCast*



Github Links







Report issues

STARS 57 | ISSUES 2 OPEN | LICENSE MIT | INSTALLATIONS 230

DOI 10.5281/ZENODO.6472890

XCast

Welcome

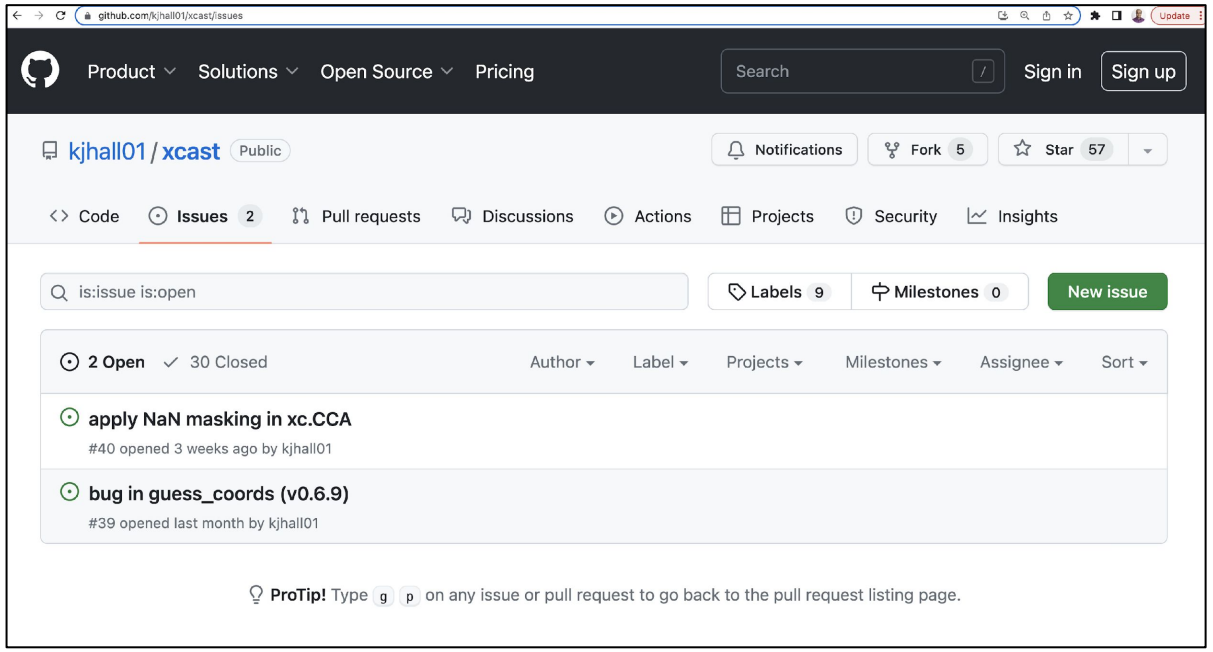
XCast is a Python Climate Forecasting toolkit - a set of flexible functions and classes that let you implement any forecasting workflow you can think of. It uses Xarray and Dask Parallelism to apply statistical and machine learning methods to any kind of gridded climate data quickly and efficiently.

Our goal is to lower the barriers to entry to innovation in climate and weather forecasting by bridging the gap between Python's gridded data utilities (Xarray, NetCDF4, etc) and its data science utilities (Scikit-Learn, Scipy, OpenCV). While XCast focuses on newer experimental techniques like quantile regression forest and extreme learning machine, it also implements many industry standard preprocessing methods and forecasting techniques from ensemble averaging to extended logistic regression. If there's something you feel is missing from XCast, have no fear- XCast is designed to be easily extensible (see BaseEstimator and @metric).

Through dask, XCast is fully compatible with many job schedulers and workload management systems. It lets you scale your machine learning-based forecasting methods to servers and institutional supercomputer clusters with ease.

XCast v0.6.9 is now live - report issues here

For more information, please check out the xcast whitepaper

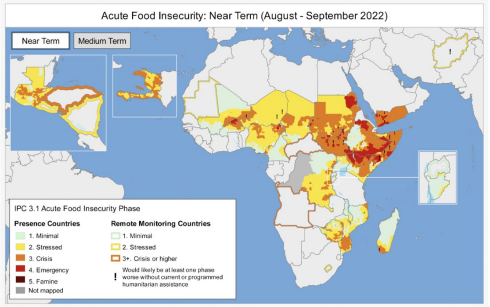


Application: Experimental Hybrid prediction system for seasonal and monthly precipitation forecasts for the FEWS NET

FEWS NET: Famine Early Warning Systems Network

FEWS NET* is a leading provider of early warning and analysis on acute food insecurity around the world.

FEWS NET provides unbiased, evidence-based analysis to governments and relief agencies who plan for and respond to humanitarian crises. FEWS NET analyses support resilience and development programming as well.

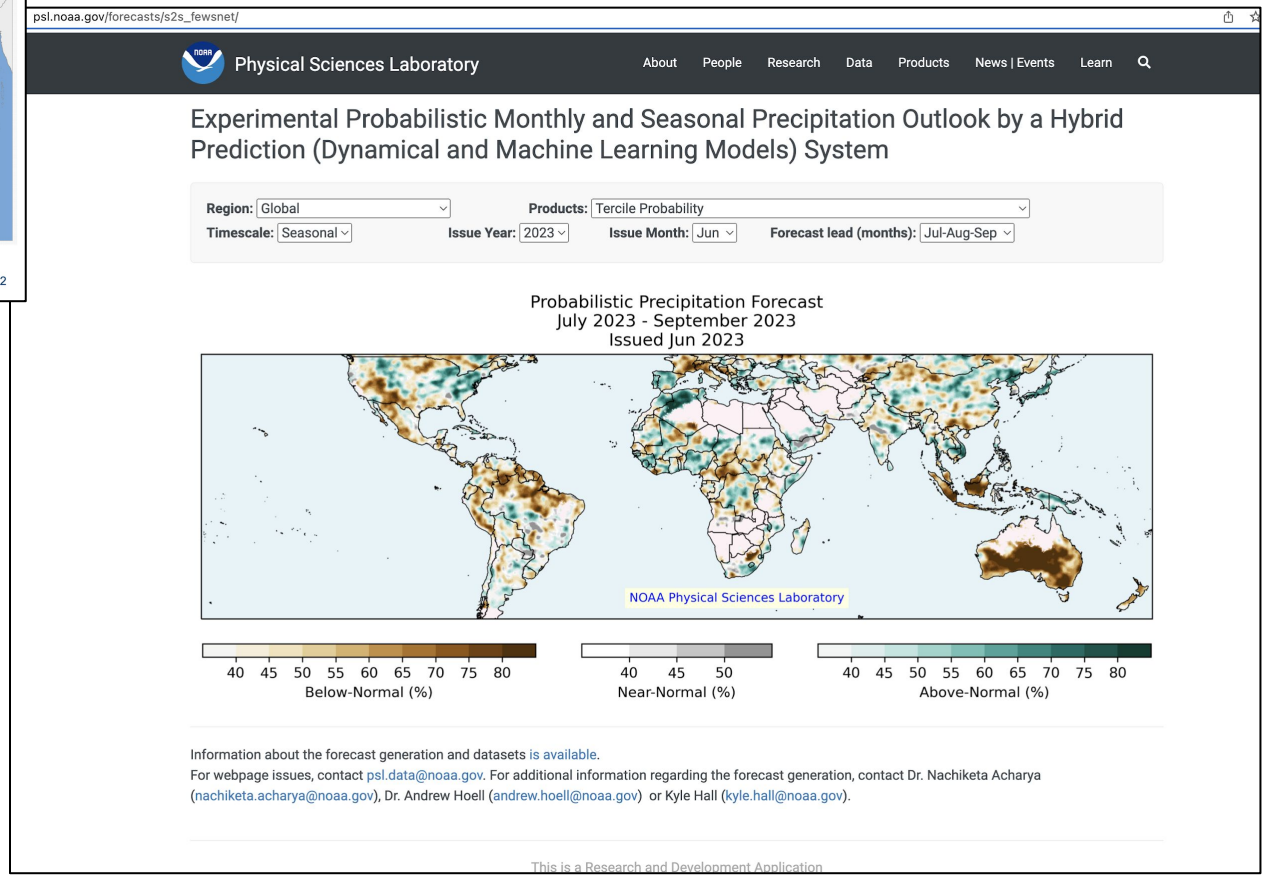


*Created in 1985 by the USAID in response to devastating famines in East and West Africa

https://fews.net/about-us

2

https://psl.noaa.gov/forecasts/s2s_fewsnet/



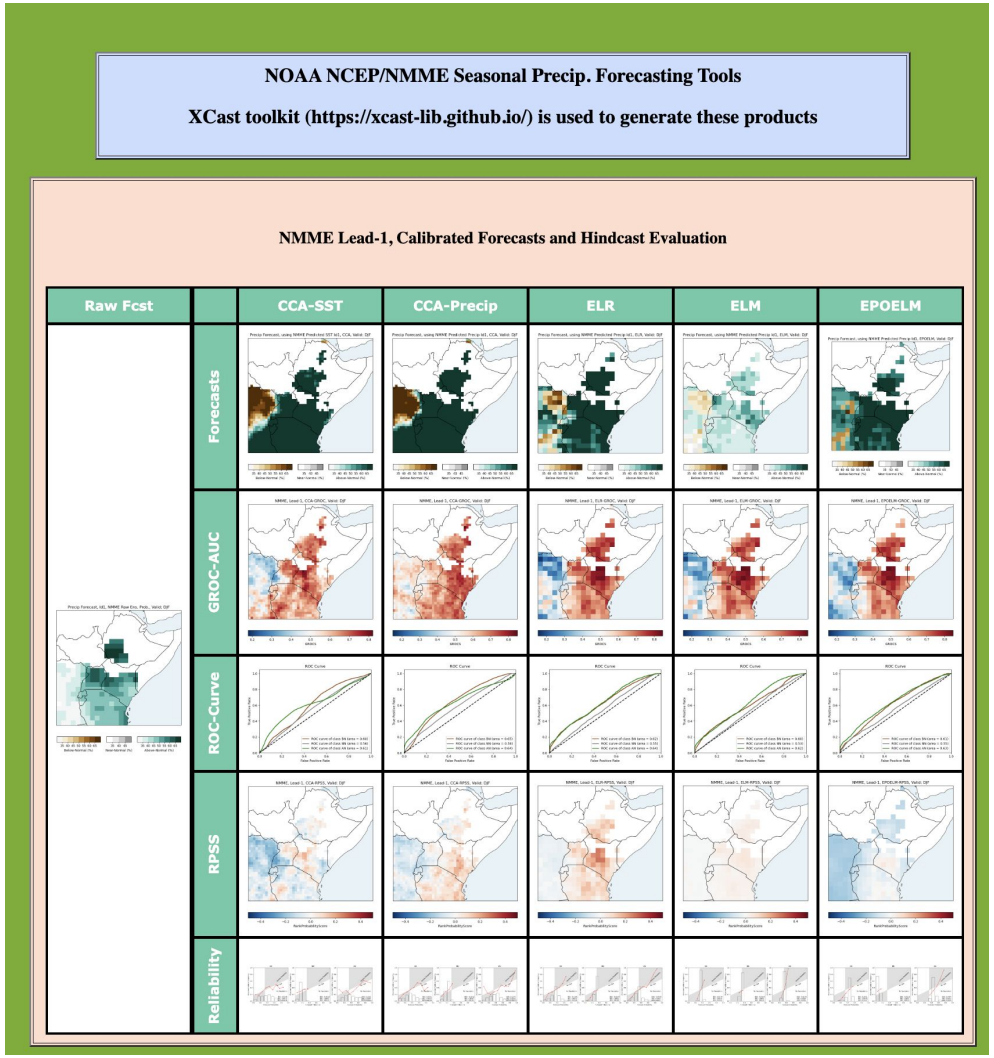
SCAN ME

EPO-ELM method used here.

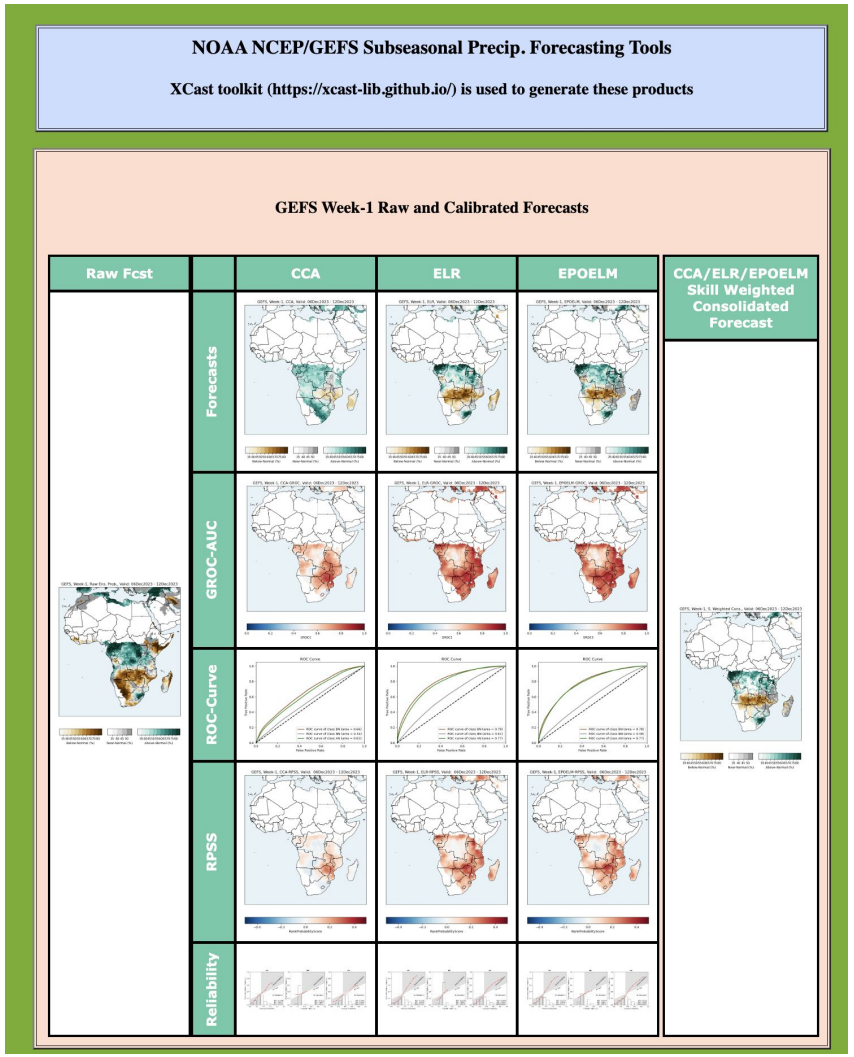
CPC's International Desk's Example

Courtesy: Endalkachew Bekele

https://ftp.cpc.ncep.noaa.gov/International/seasonal/EAfr/

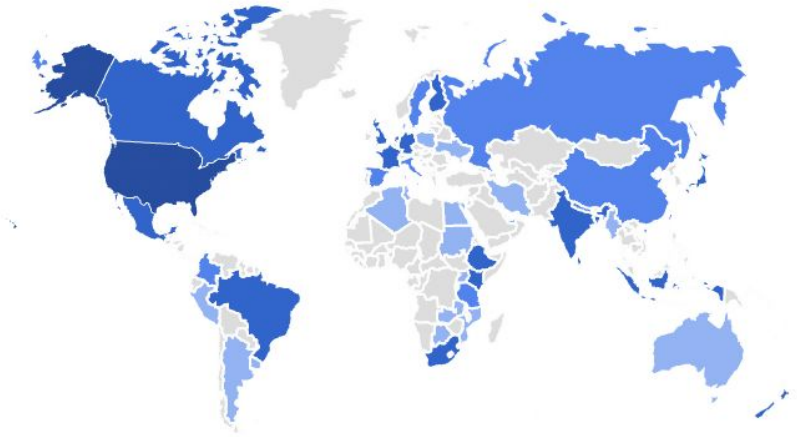


https://ftp.cpc.ncep.noaa.gov/International/wk34/xcast/



Latest Google Analytics Report for the *XCast* webpage (March,2024)

Users by Country ID



COUNTRY		USERS
United States	345	↑832....
India	36	↑620....
South Africa	18	↑350....
Bangladesh	10	↑900....
Germany	10	↑400....
Kenya	10	↓33.3%
Sri Lanka	7	-

Last 90 days

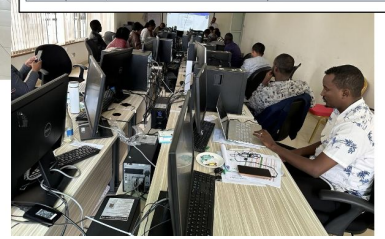
View countries →

XCast is used by many users all over the world.

XCast Trainings

XCast Training at ICPAC, Nairobi (Nov,2024)

	Participants	Country/Organization
1	Ferdinand Niyomukunzi	Burundi
2	Houssein Farah Adawe	Djibouti
3	Hailu Negese Kabebe	Ethiopia
4	Christine Maswi	Kenya
5	Sandrine Guhirwa	Rwanda
6	Abdisamad Mohamed Ali	Somalia
7	Dolly Elviegi Faustino	South Sudan
8	Mohammed Seif Aldeen Saeed Ebaid Allah	Sudan
9	David Kisira	Uganda
10	Yimer Assefa	Ethiopia
11	Samuel Thon	South Sudan
12	Emmanuel Qurino	South Sudan
13	Patricia Okau	Uganda
14	Saeed Yagoub	Sudan
15	Otim Faustine Charles Obeke	Uganda
16	Dr. Titike Bahaga	ICPAC
17	Dr. Hussen Seid	ICPAC
18	Dr. Masilin Gudoshava	ICPAC
19	Eunice Koeach	ICPAC
20	Anthony Mwanthi	ICPAC



XCast Training for Sri Lankan Met. And Indian Met. Officers, virtually (Nov,2024)

S. No.	Name	Organization	Email ID
1	Ms. Anusha Rashanthi Patabadi Warnasooriya.	Department of Meteorology Sri Lanka	rashanthie@yahoo.com
2	Ms. Asurappulige Dulari Gangani Fernando.	Department of Meteorology Sri Lanka	dularifdo@yahoo.com
3	Mr. Jayasinghe Sepalage Dharshana Shamil Premathilake.	Department of Meteorology Sri Lanka	darshana.shamil@gmail.com
4	Ms. Kankani Achchi Kankanamalage Tilanthi Wimal Weerasinghe.	Department of Meteorology Sri Lanka	twimal@yahoo.com.au
5	Ms. Tanu Sharma	India Meteorological Department	tanu35bhp@gmail.com
6	Ms. Sandhya Jose	India Meteorological Department	sandhya.kalluvilayil@gmail.com
7	Mr. Arvind Pagire	India Meteorological Department	arvindpagire1995@gmail.com
8	Ms. Jasmine Takle	India Meteorological Department	jas.takle@gmail.com
9	Ms. Madhuri Musale	India Meteorological Department	mmusale774@gmail.com

XCast training is part of WMO's regional Climate Outlook Forums: SASCOF, ASEANCOF, GHACOF etc.

XCast in S2S Newsletter

Subseasonal-to-Seasonal S2S Prediction Project

Newsletter No. 21

WCRP World Climate Research Programme

WWRP World Weather Research Programme

In This Issue…

News:

Topics:

1. What is S2S?
2. Six sub-projects in S2S Phase II
3. Upcoming Events

1. What is S2S ?

To bridge the gap between medium-range weather forecasts and seasonal forecasts, the World Weather Research Programme (WWRP) and World Climate Research Programme (WCRP) launched a joint research initiative in 2013, the Sub-seasonal to Seasonal Prediction Project (S2S). The main goal of this project is to improve forecast skill and understanding of the subseasonal to seasonal time-scale, and to promote its uptake by operational centres and exploitation by the applications communities.

Phase II of the S2S project began in January 2019 and will continue until 2023. A new set of scientific sub-projects has been developed, as outlined in the sidebar in next pages. Enhancements to the database will be made including access to the S2S ocean and additional models. The second phase will also include new research-to-operations activities and a real-time supplications initiative introduced in this edition of the newsletter.

S2S Phase II Proposal is available at http://s2sprediction.net/file/documents_reports/P2_Pro.pdf

WWRP Symposium for its New Implementation Plan 2024-2027

Munehiko Yamaguchi

Scientific Officer for the World Weather Research Programme at WMO

***Fig. 1**. Group photo taken during the WWRP Symposium at WMO HQ Attic.*

The World Weather Research Programme (WWRP) Symposium was successfully held in a hybrid format from 22 – 26 August 2022 in Geneva, Switzerland (WMO HQ). The current Implementation Plan of WWRP is coming to an end in 2023 and we are working together towards a new plan for 2024-2027, aligning with the WMO Strategy for the same period. The Symposium was attended by the WWRP Scientific Steering Committee (SSC), WWRP Working Group (WG) and Expert Team (ET) co-chairs, as well as leaders of WWRP-related projects and invited speakers (see fig. 1). During the first half of the week– we had presentations by invited speakers discussing the plans for our future related to the draft Implementation Plan of WWRP. The second half of the week was dedicated to the annual meeting of the WWRP SSC which included feedback and plans from all WGs, ET, projects and our Research and Development Projects/Forecast Demonstration Projects (RDPs/FDPs). The program of the Symposium and all presentation materials are available on the Symposium website.

An Introduction to XCast: A Python toolkit for AI/ML based S2S forecasting

Nachiketa Acharya[1,2] and Kyle Joseph Chen Hall[1,2]

[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Boulder, CO, United States, [2]National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory, Boulder, CO, United States

Motivation and Aim:

Sub-Seasonal (S2S) climate forecasts suffer from a significant lack of prediction skill at lead times longer than two weeks. Over the past decade, there have been intense research efforts within the World Weather Research Programme–World Climate Research Programme (WWRP/WCRP) Sub-seasonal-to-Seasonal (S2S) Prediction project. These efforts largely have been aimed at improving predictions by making the output of state-of-the-art global coupled ocean-atmosphere circulation models (GCM) available in public databases like the S2S and SubX projects. However, these ensembles forecast products cannot be used directly, and require further calibration.

While the traditional statistical bias correction of GCM offers some measure of skill, there is significant interest in the capacity of Artificial Intelligence (AI)/Machine Learning (ML) based calibration approaches to improve S2S forecasts. In order to improve S2S precipitation and temperature forecasts at week 3-4 lead times with AI/ML,

***XCast* will be feature in upcoming WMO guidance document for sub-seasonal forecasting.**

Concluding Remarks

- *XCast* build objective forecasts using multiple dynamical model output predictors with enforcing AI/ML based calibration produces both deterministic and probabilistic forecasts after hindcast skill assessment for S2S scale.

- *XCast*, a user friendly tool, can run in laptop to cluster/cloud with parallel computing process.

- "It provides a platform for scientists to perform and compare several post-processing/calibration methods"- Endalkachew Bekele, NOAA-CPC

Future Plans

- Proper funding to develop and maintain *XCast*.
- Make a community of users and developers.
- conda to pip install for using in Google Colab and similar platforms.

Thank you!

Any Questions/future collaborations?

dr.nachiketaacharya@gmail.com/nachiketa.acharya@noaa.gov