



A Continental US testbed for Basin-scale S2S Climate Predictions supporting water management

Andy Wood^{1,2} and Josh Sturtevant^{2,3}

¹NCAR Climate and Global Dynamics – Boulder, Colorado

²Colorado School of Mines, Civil & Env. Engineering – Golden, Colorado

³Lynker – Boulder, Colorado

HEPEX Workshop: Forecasting across spatial scales and time horizons

13-15th September 2023

SMHI, Norrköping, Sweden



Gap, Motivation

Sub-seasonal to seasonal (S2S) climate prediction is fundamental to myriad, diverse applications across many sectors.

- the ability to predict climate beyond weather scales (2 weeks to 9 months) can inform water, energy, agriculture, terrestrial and aquatic ecosystems, infrastructure development, shipping and navigation, ...

There is **no coherent source of guidance** on the '*latest and greatest*' S2S climate information for any particular application

- stakeholders with particular S2S climate needs
- S2S forecast developers (such as CGD) seeking a benchmark for their own modeling capabilities

Some raw model forecast skill assessments exist (in literature, on provider websites)

- default grids, default predictands
- not comprehensively intercomparable
- lacking breadth: e.g., only dynamical, only raw (no post-processing)



Example needs from current projects

Colorado River management needs

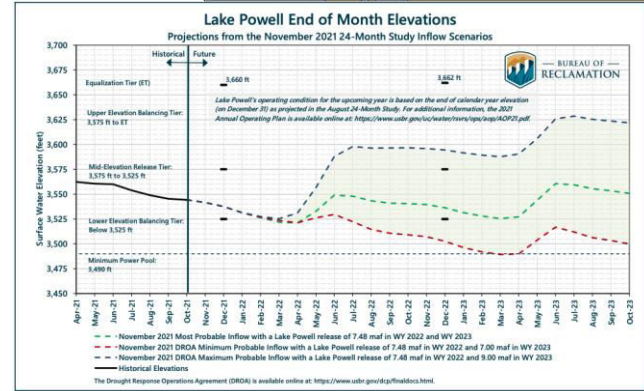
- Operated on predictions: short-range, S2S, 2 year ('24-month study), and 5 year outlooks
- Almost no use of S2S climate predictions, despite intense interest
 - example of current approaches: 5-year average climatology for certain inflows; elsewhere unconditional ESP after 10 days
- Currently looking into North American monsoon season S2S predictions with the management agency, Reclamation



Hoover Dam / Lake Mead



Lake Powell

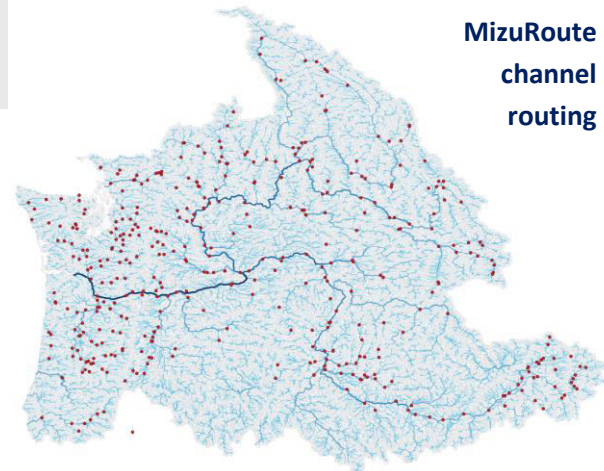
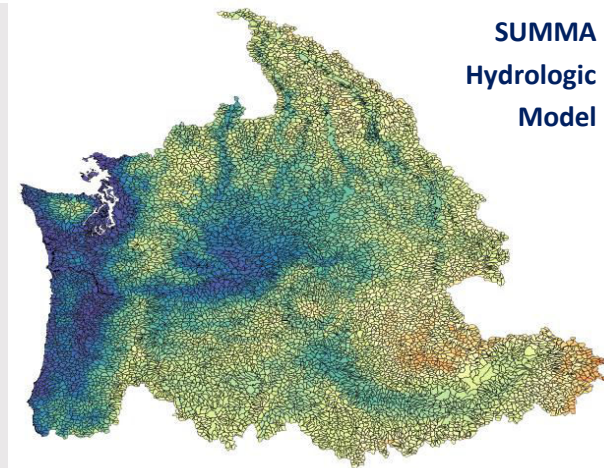
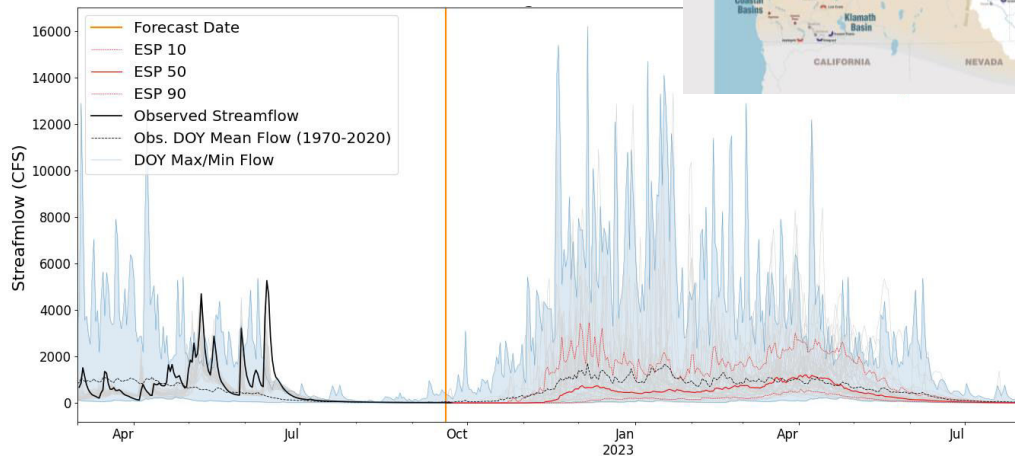


Example needs from current projects

Columbia River system management

- USACE & Reclamation agencies
- Need S2S climate predictions to drive inflow forecasts
- Using GEFS for weather (1-7 day)
- GEFS for S2S (weeks 2-4)
- NMME for seasonal (Months 2-4)

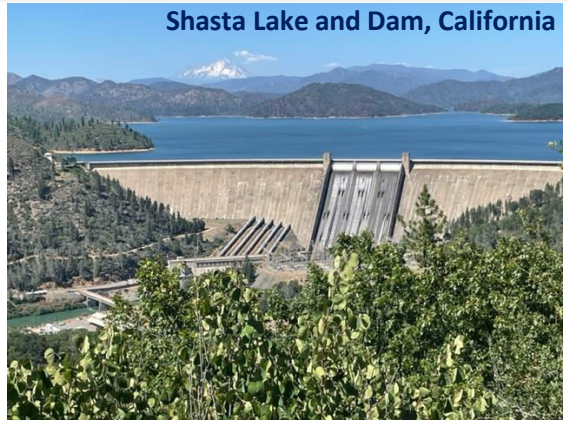
Ensemble S2S streamflow prediction



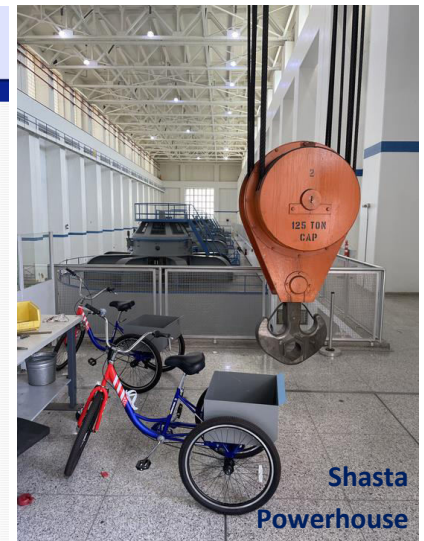
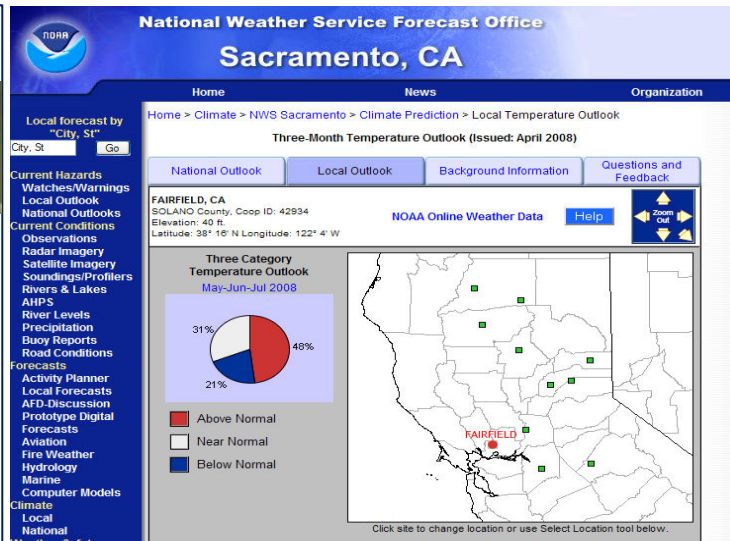
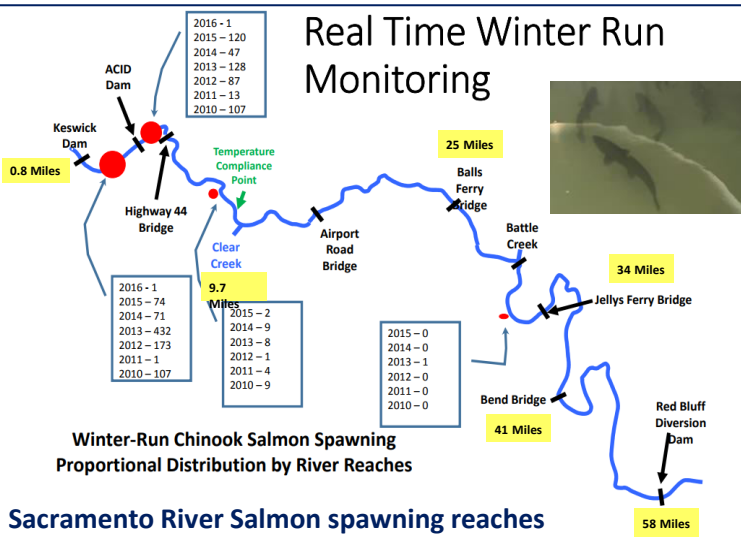
Example needs from current projects

Sacramento River system management

- Reclamation releases and use of cold water storage is determined by S2S stream temperature predictions from March to November
- Current approach: tercile forecasts of temperature + conditional resampling of met. inputs to a stream temperature model
- It has marginal skill in the first month only



Tercile climate forecasts from CPC



Testbed concept to serve multiple projects over time

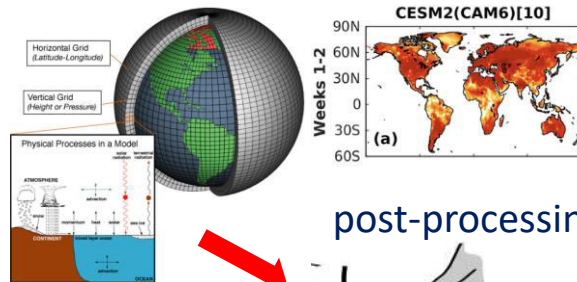
A testbed should merge cross-cutting expertise

- climate forecast development
- data-driven post-processing
- stakeholder applications
- geospatial data analysis
- interactive visualization

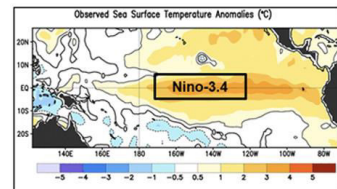
It should include

- stakeholder predictands & engagement
- broad intercomparison & benchmarking
- web focal point as community resource

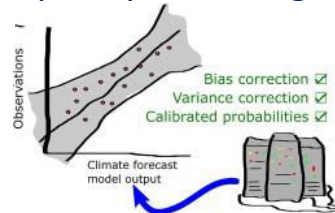
dynamical forecasts



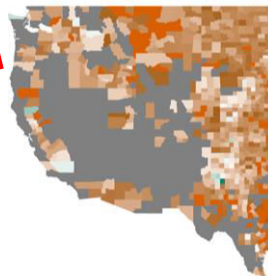
empirical forecasts



post-processing

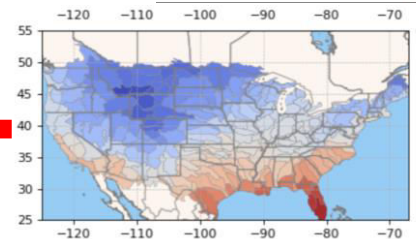
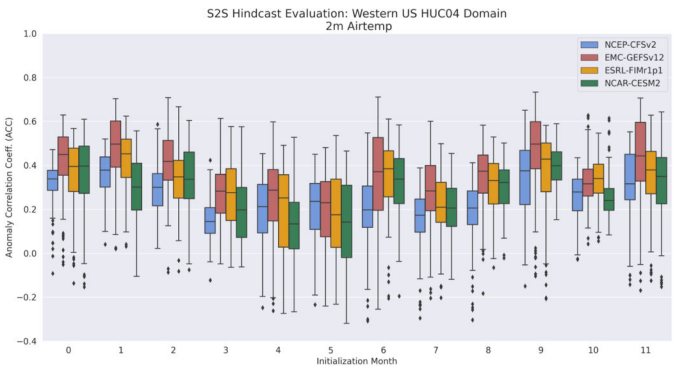


sector specific predictands

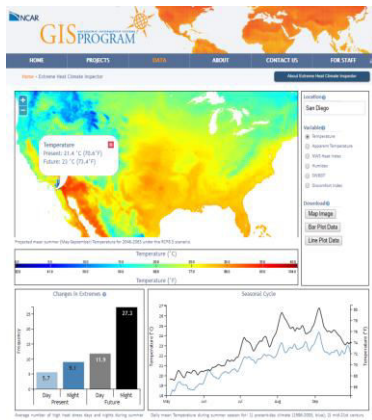


feedback to developers

verification and benchmarking



online interactive visualization



Strategy for S2S climate forecast use

Hydrologic prediction needs seamless forecast meteorology

- short (to medium) range weather forecasts
- sub-seasonal (eg weeks 2-4)
- seasonal (months 2+) climate signals

(1) Handle weather scale prediction relatively directly

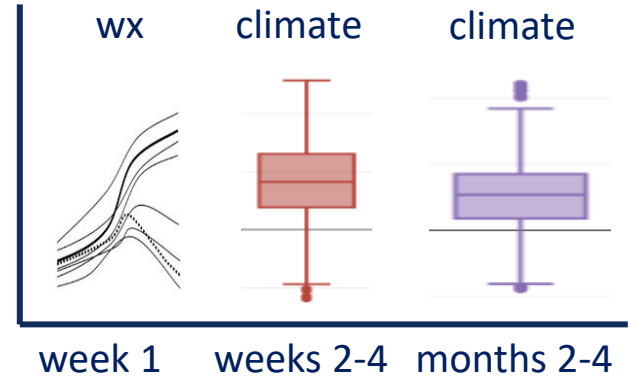
- bias-corrected direct output from NWP

(2) Handle climate-scale prediction indirectly

- climate-forecast conditioned weather generation

Pros & cons for different sequencing

- optimal predictability varies, but expediency matters



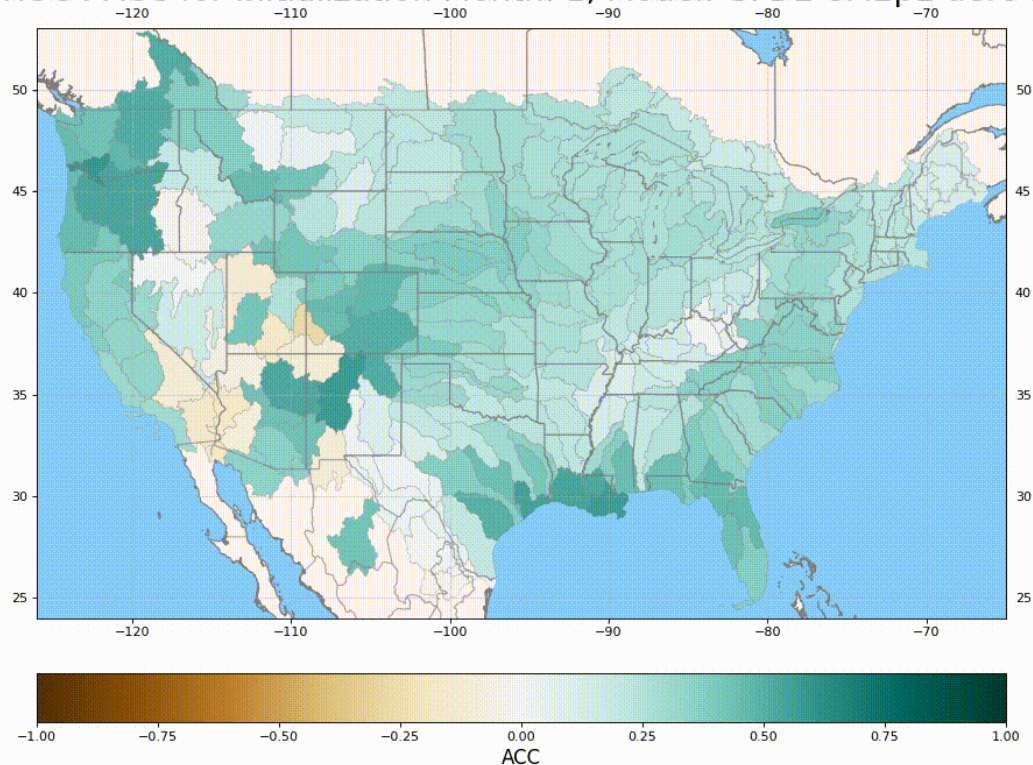
Examples of testbed S2S climate forecast investigation

Individual model skill evaluations

- Example: one dynamical model (GFDL) from NMME
- From various sources, we have not found a great solution for months 2/3 precipitation or air temperature (or month 2)

Note: have yet to add ECWMF's forecasts to testbed

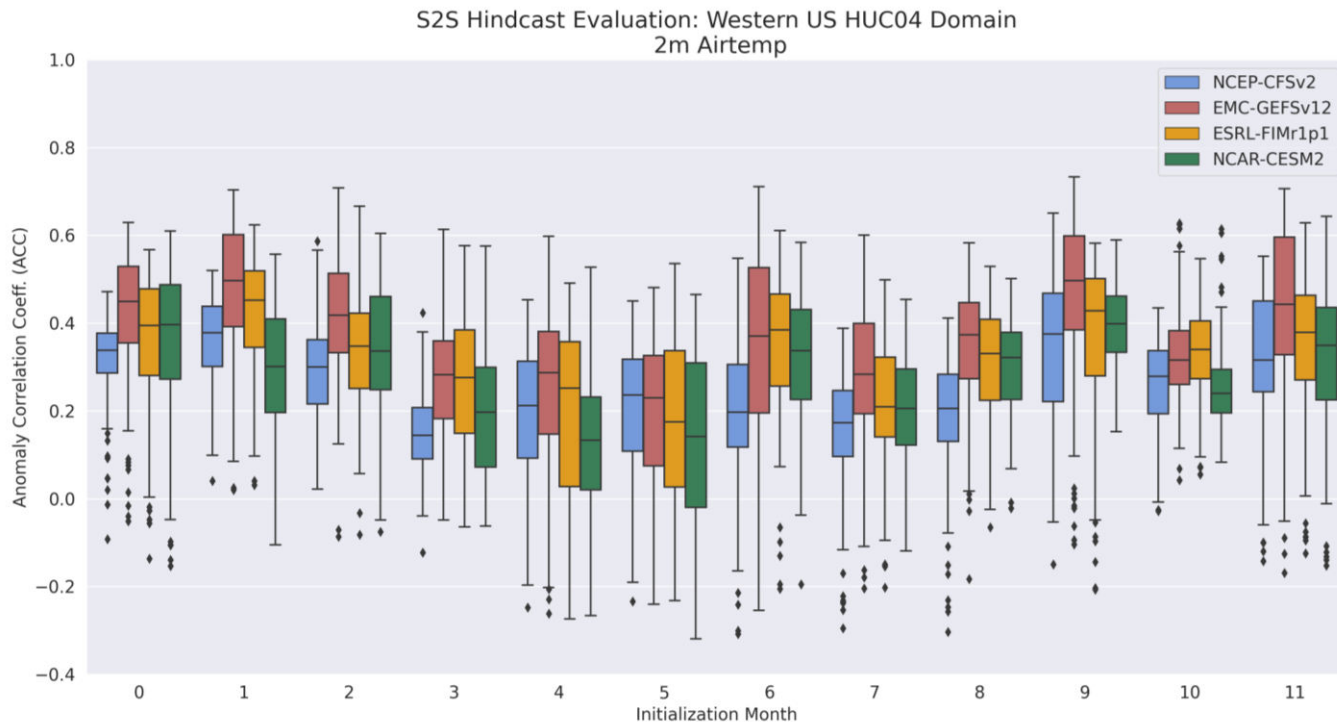
NMME Hindcast Evaluation (1982-2010) of Months 2-3 Airtemp HUC4 ACC for Initialization Month: 1, Model: GFDL-CM2p1-aer04



Examples of testbed S2S climate forecast investigation

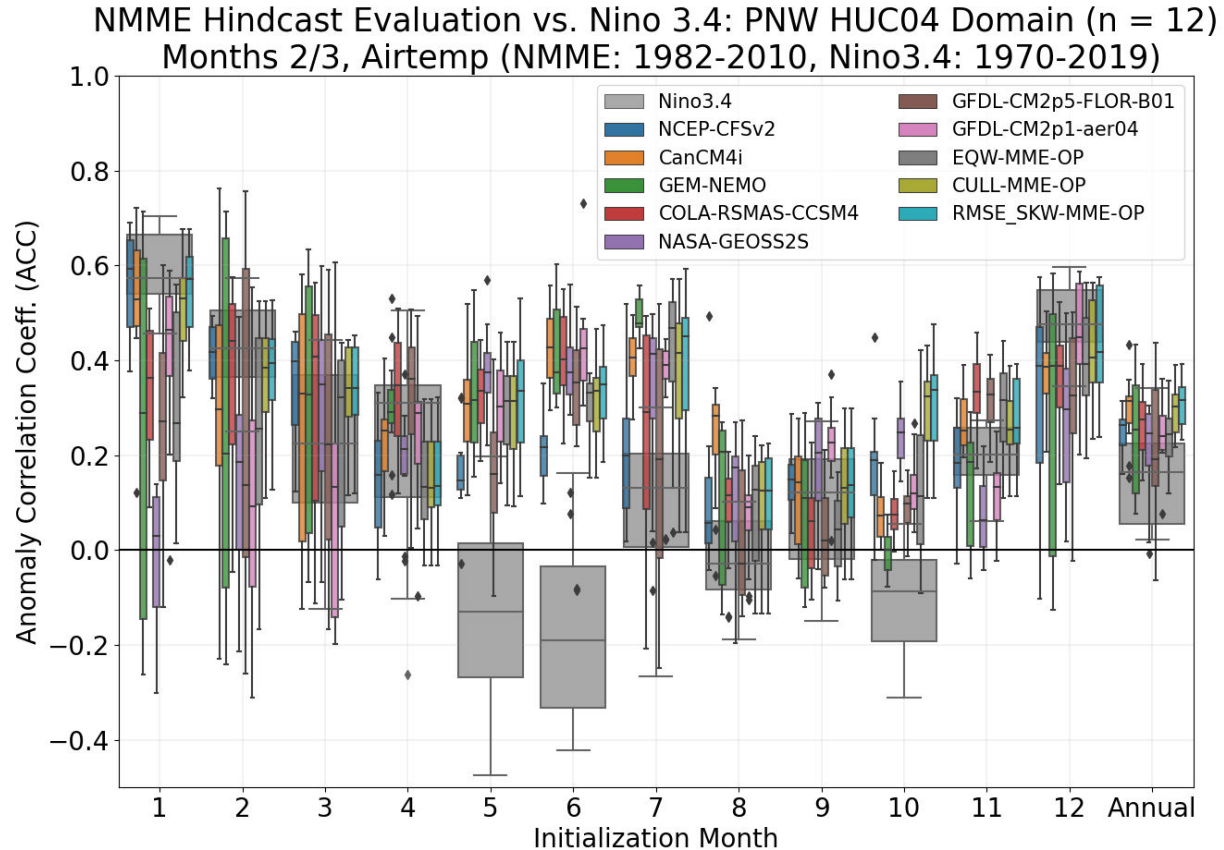
Comparative dynamical model skill evaluations

- Weeks 2-4 temperature and precipitation
- We looked mainly at the Sub-X models (dynamical sub-seasonal forecasting)
- To date, the NCEP Global Ensemble Forecast System (GEFS) provides the highest skill



Example: comparing empirical, dynamical predictions, some post-processing

- Simple empirical predictions (index-based) are still an important benchmark with dynamical and hybrid models (example, Nino3.4)
- Simple post-processing can help
- Months 2/3 skill is often too low in Western US; will focus on 'season 2' instead – eg, months 2-4, 3-5 ...
- There are more options to explore with the testbed methods (e.g., ML)



Exploring ML-based post-processing (hybrid forecasts)

- e.g., XCast
- Used in IRI calibrated products
- Took 3rd place in AI/S2S Challenge
- We're collaborating with the XCast developer at NOAA/ESRL
- Other ML tools/packages are of interest

XCast

1. About

2. Installing XCast

3. Data in XCast

4. Citing XCast

ACPAC

BaseEstimator

CCA

CrossValidator

EOF

EmpiricalTransformer

Ensemble

GammaTransformer

LeaveOneYearOut

MLR

MinMax

Normal

OneHotEncoder

Search XCast

STARS

60

ISSUES

2 OPEN

LICENSE

MIT

INSTALLAT

DOI

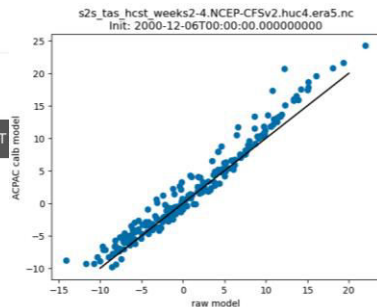
10.5281/ZENODO.6472890



Welcome

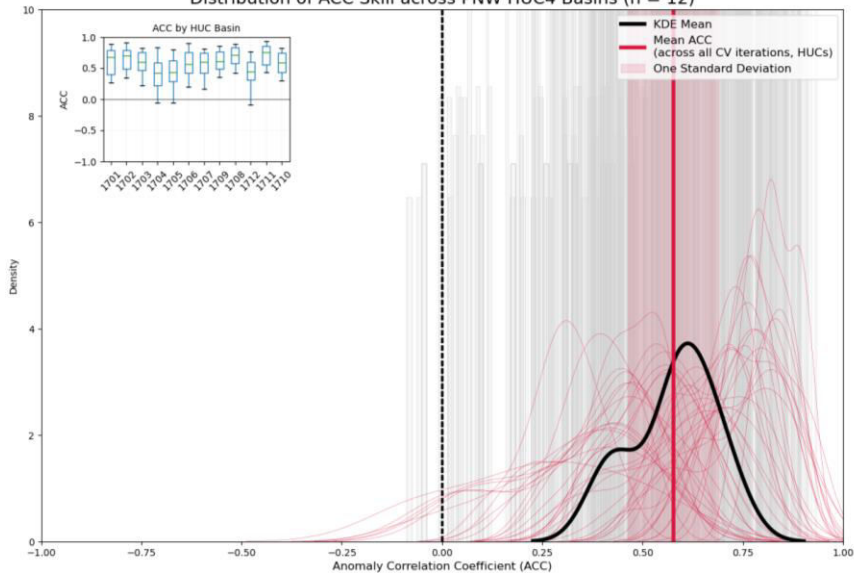
XCast is a Python Climate Forecasting toolkit – a set of flexible functions and classes that let you implement any forecasting workflow you can think of. It uses [Xarray](#) and [Dask Parallelism](#) to apply statistical and machine learning methods to any kind of gridded climate data quickly and efficiently.

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).

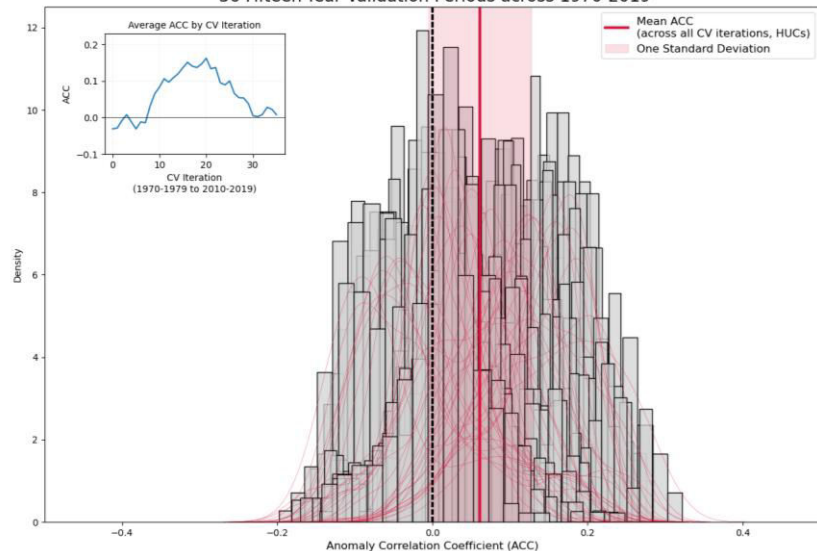


The challenge of small samples and skill metric uncertainty

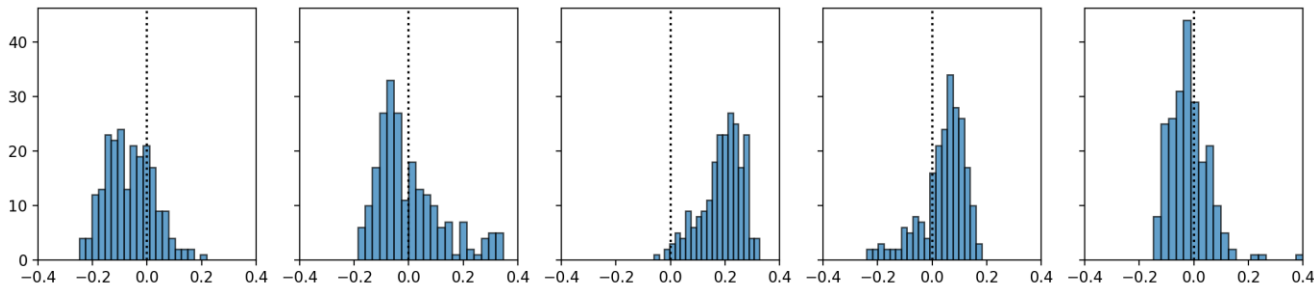
January Nino3.4 MLR for Seasonal (FMA) Airtemp Lagged 1-Month
Distribution of ACC Skill across PNW HUC4 Basins (n = 12)



Nino3.4 PDO_NPGO MLR for 0-Month Airtemp
Distribution of ACC Skill across CONUS HUC4 Basins (n = 222)
36 Fifteen-Year Validation Periods across 1970-2019



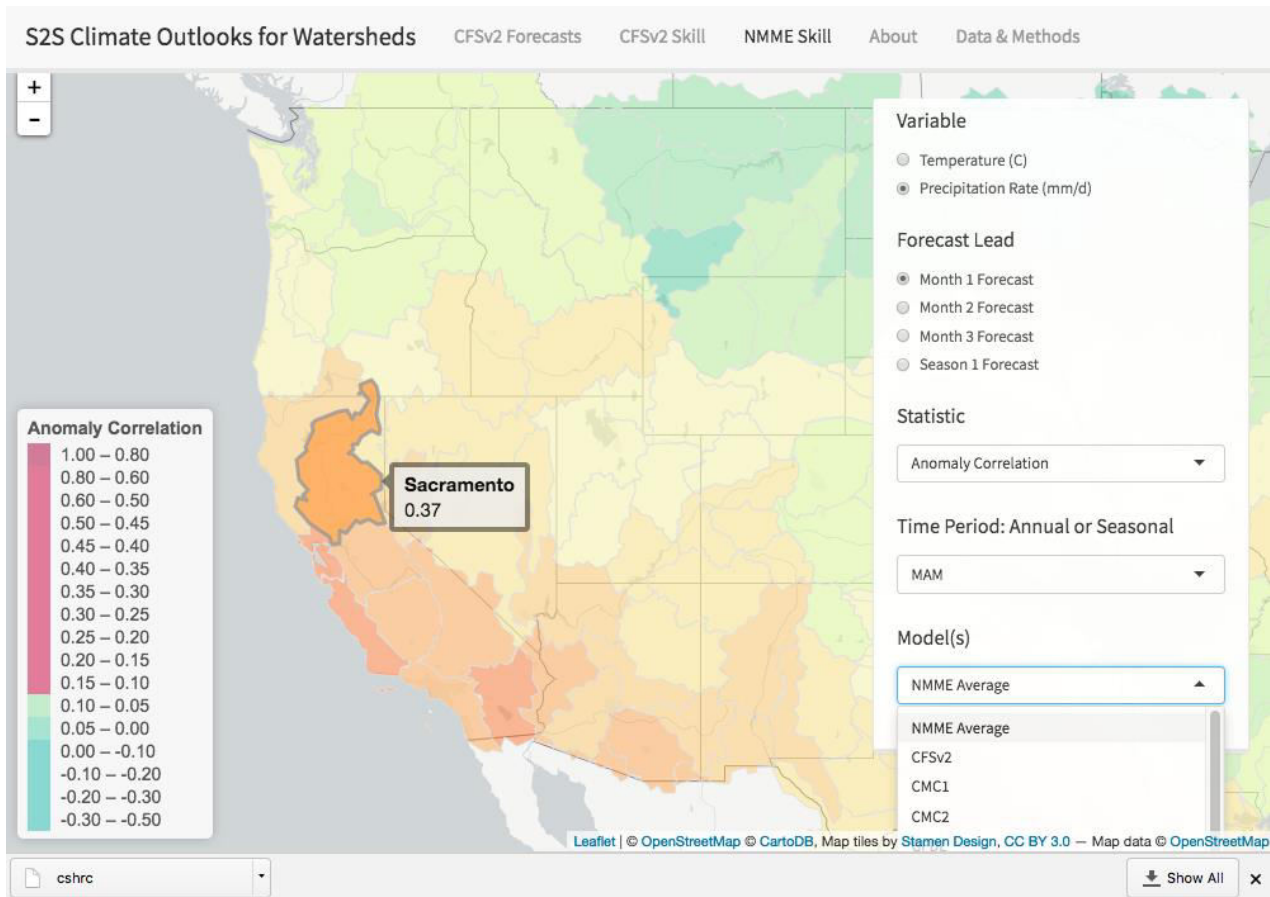
0-Month Airtemp ACC Distribution for Nino3.4/NPGO MLR under a 5-fold Cross Validation (n = 222 HUC4s)



Interactive visualization

A previous project engaged water stakeholders via a watershed focused climate evaluation

A next step will be to revive this platform with the new data from the Testbed effort



Summary

- To support multiple forecast projects that need S2S climate predictions, we've created a testbed focusing on basin-level climate forecast skill
- The testbed shares a common infrastructure for evaluating dynamical, empirical, and hybrid predictions – including ML approaches
- We're assessing different convenient predictand periods
 - week 2, weeks 2-3, weeks 2-4, months 2, 3, ..., months 2-3, season 1, 2, 3 ...
- The testbed should help to benchmark new, individual efforts as they appear
- Funding and collaboration is currently somewhat ad hoc – support is from several different projects versus one dedicated coherent whole
- Estimating skill based on small samples and short records requires caution

Contact: andywood@ucar.edu, NCAR; awwood@mines.edu, CS Mines

