Center for Western Weather and Water Extremes
SCRIPPS INSTITUTION OF OCEANOGRAPHY AT UC SAN DIEGO

S2S PREDICTION OF WINTER PRECIPITATION: EXPLORING NOVEL STATISTICAL AND MACHINE LEARNING METHODS

Western States Water Council Meeting
May 18, 2022

NASA Earth Science Applications: Water Resources Project (2022-25)

Agniv Sengupta (CW3E, Scripps Institution of Oceanography),
Bin Guan (PI: UCLA), Duane E. Waliser (JPL), Jeanine Jones (California DWR)

UC San Diego | SCRIPPS INSTITUTION OF OCEANOGRAPHY

Applied Science Questions

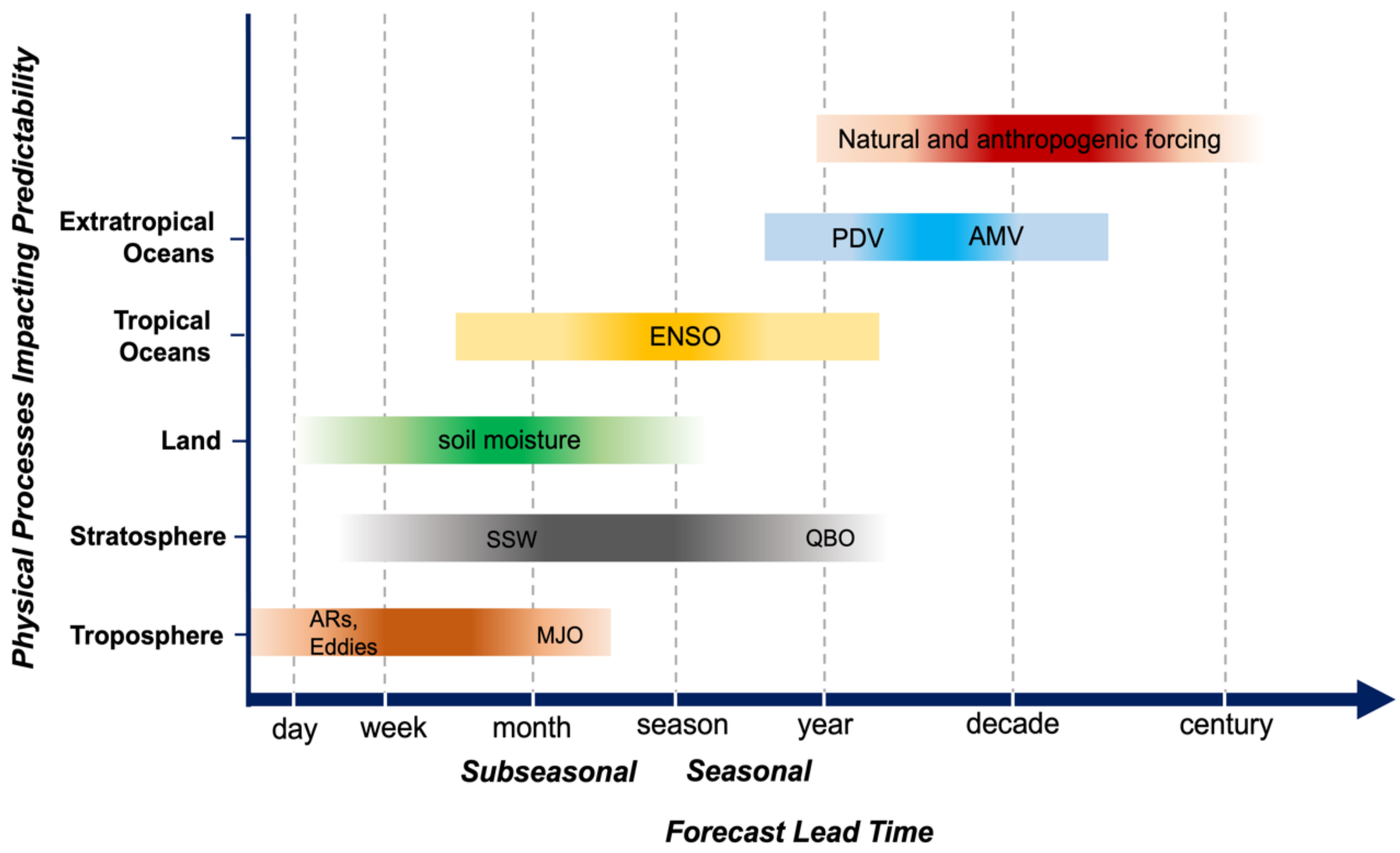


Figure 1. Schematic representation of physical processes in the Earth system governing predictability of precipitation across a range of temporal scales (weather, subseasonal, seasonal, annual, and decadal-multidecadal).

Key Questions

- How can we optimize the predictive influence of key predictors of winter precipitation in the western U.S.?
- What are the optimal temporal lags of these predictors for skillful S2S prediction?
- How do we optimally combine observations with dynamical model outputs for improving precipitation forecasts within a hybrid framework?

Common Challenges in Current Forecast Systems

Prediction Skill: Winter Precipitation

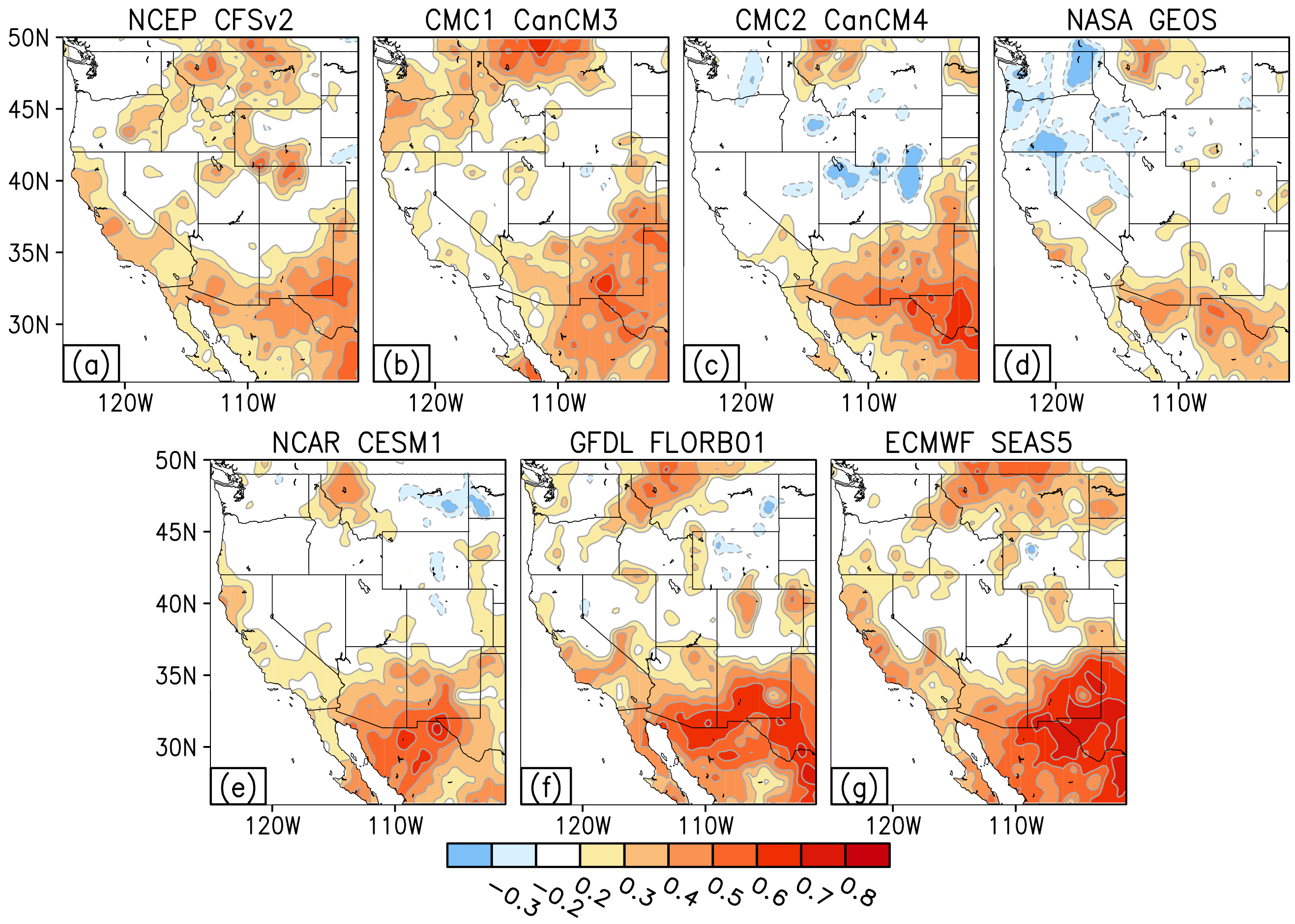


Figure 2. Seasonal prediction skill of winter precipitation in leading dynamical models from the North American Multi-Model Ensemble (NMME) project and ECMWF, displayed as correlations between model hindcast and observed precipitation anomalies. Period of analysis: 1982–2010 winters (DJF).

- Consistently low skill of precipitation forecasts across the western U.S. after a 2-week lead-time (Slater et al. 2019)
- Very little improvement for seasonal forecasts of precipitation across successive upgrades of NMME model versions (Becker et al. 2020)

Common Challenges in Current Forecast Systems

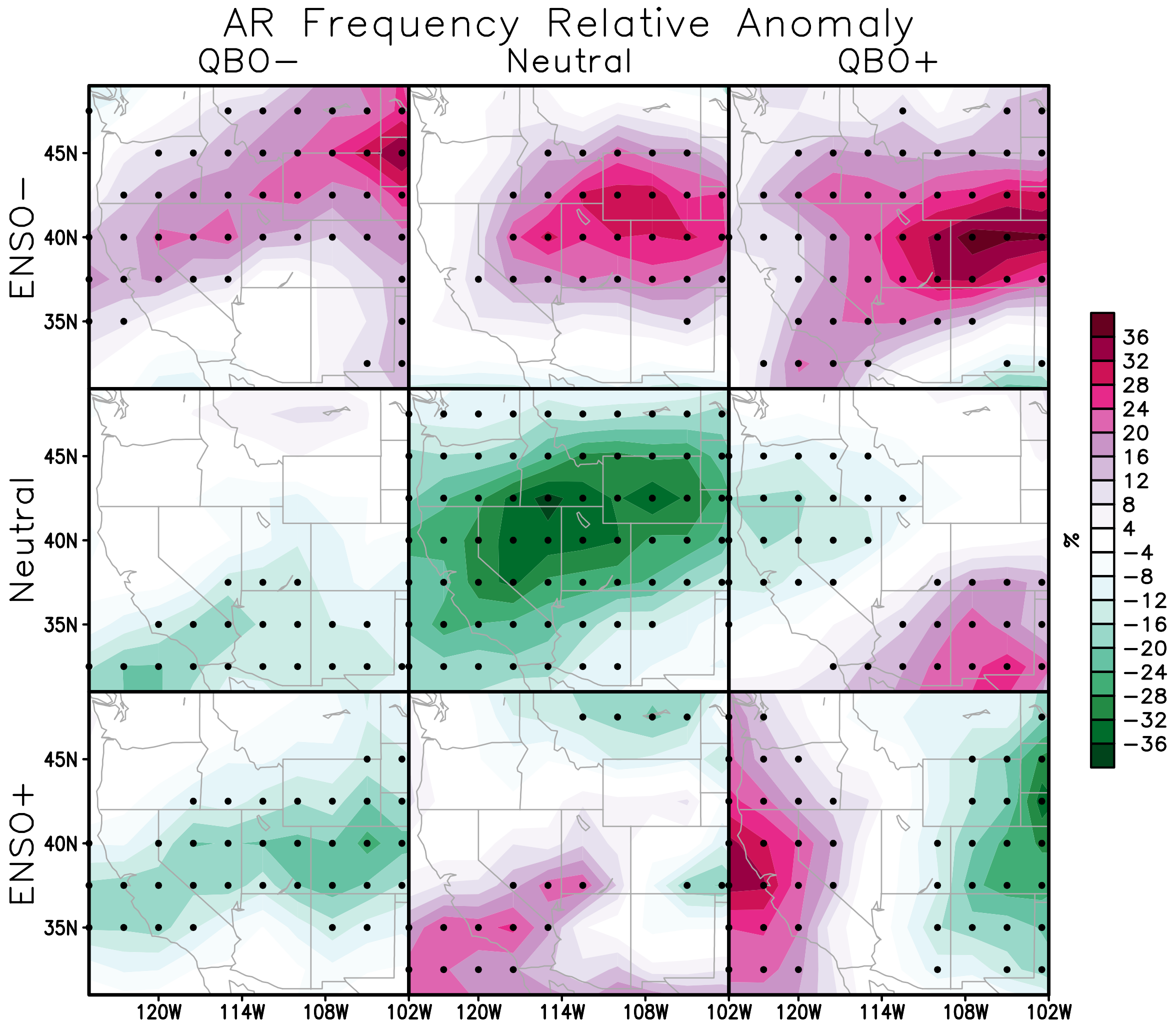


Figure 3. Example of non-linear relationship between ENSO, QBO, and ARs. Shaded are AR frequency anomalies normalized by the seasonal mean. The result illustrates the asymmetric modulation of AR frequency between opposite phases of ENSO (QBO) during a given phase of QBO (ENSO), suggesting non-linear interactions.

- Failure to consider the non-linearity in process interactions, e.g., in statistical forecasts which historically rely on the use of lagged correlations
- Limited size of observational data leading to overfitting when trying to accommodate multiple sources of predictability and their related interactions

Figure courtesy: Bin Guan

Proposed Application & Objectives

- Proposed application: Develop experimental methods for prediction of winter precipitation considering multiple predictors (e.g., sea-surface temperature, outgoing longwave radiation, stratospheric state, soil moisture, etc.) using statistical and machine learning methods
- Advantages:
  - Ability to accommodate *multiple sources of predictability* in a unified framework
  - Extracts clearly *delineated predictors* which are mutually orthogonal by construction
  - Ability to model non-linear process interactions
  - Less prone to overfitting

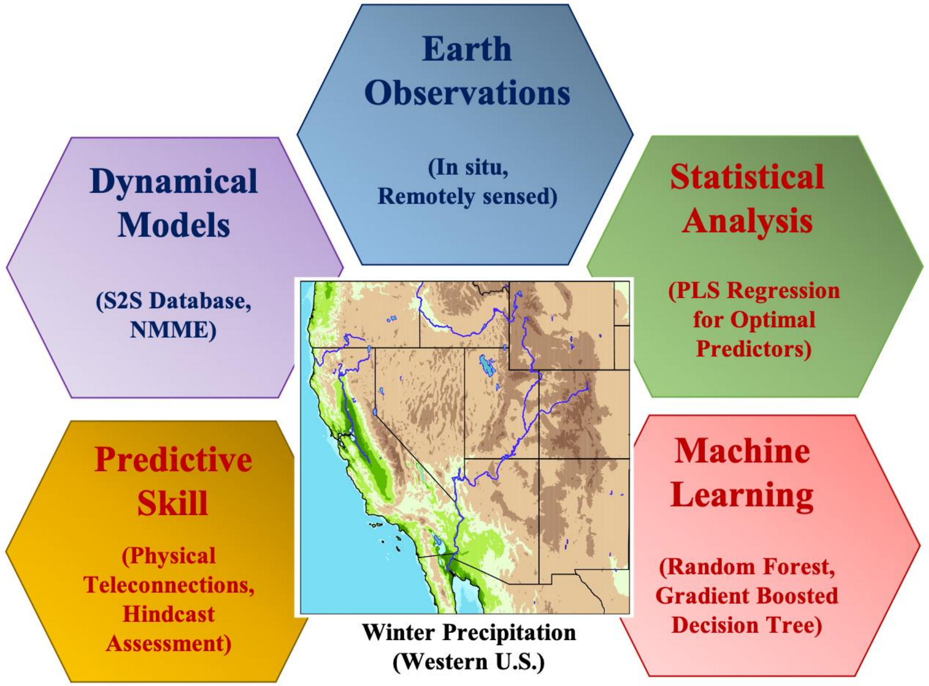


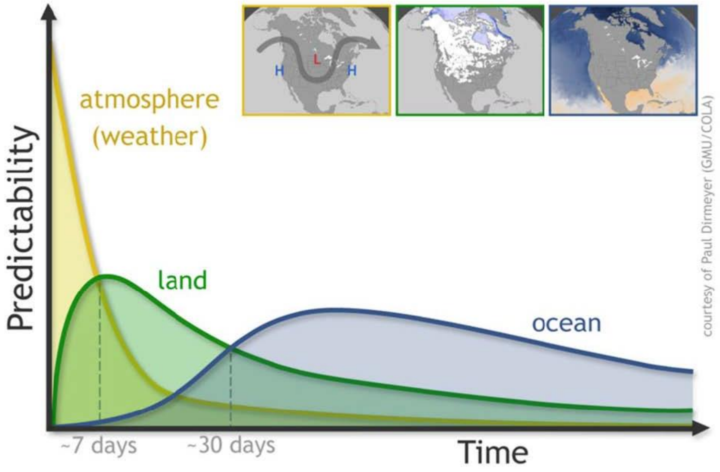
Figure 4. Graphical schematic displaying the focus areas of the proposed work

- Objectives:
  - Characterize the key sources of predictability of winter precipitation of relevance for drought response and water management in the western United States
  - Design, validate, and implement an experimental prediction tool to optimally combine the predictors and generate forecasts of winter precipitation
  - Evaluate the representation of key S2S processes and phenomena in current weather/climate prediction models, assess model strengths and shortcomings.

Datasets and Analysis Methods

❑ Data and modeling resources:

- Observations/Reanalyses: *Sea-surface temperature* (HadISST, OISSTv2); *Geopotential heights and winds* (MERRA-2, ERA5); *QBO* (NOAA PSL); *Outgoing longwave radiation* (NOAA Interpolated OLR); *Soil moisture* (SMAP, ESA CCI); *Precipitation* (NOAA CPC-Unified, GPM-IMERG, PRISM)
- Dynamical models: WWRP/WCRP S2S Database (subseasonal); NMME (seasonal)
- Large Ensembles of Climate Models: NCAR Community Earth System Model Large Ensemble Project



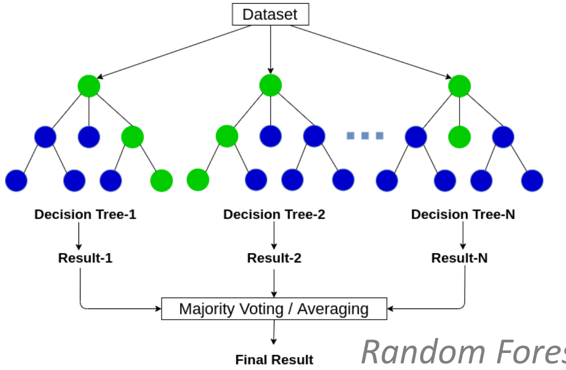
❑ Proposed Methods:

(1A) Extract the predictor variables at optimal lags from statistical analyses specifically designed to accommodate multiple predictive sources (in the atmosphere, ocean, and land) without overfitting

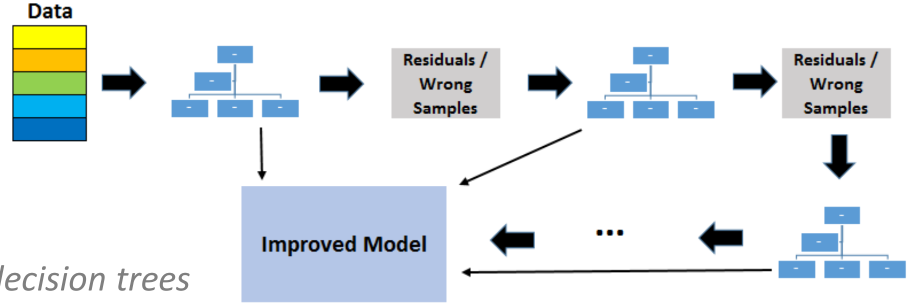
(1B) Apply ML algorithms on the extracted predictors and build an experimental predictive model by additionally accommodating non-linear process interactions

(2A) Hybrid approach: Directly apply supervised and unsupervised ML algorithms in model space on large ensembles of climate models

(2B) Implement *transfer learning* method by updating the pre-trained weights on observations



Random Forest



Gradient-boosted decision trees

Extraction of Predictors: *A prototype analysis*

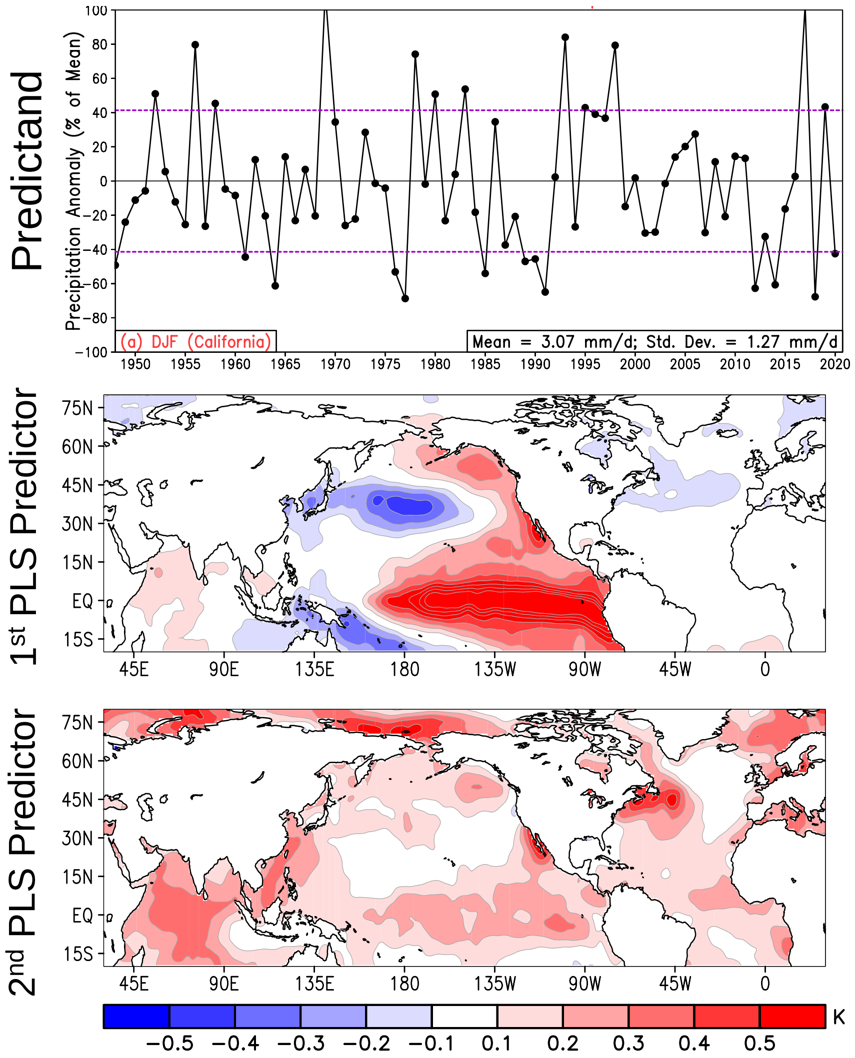


Figure 6. (upper) Dec–Feb CA precipitation anomaly expressed as % of long-term mean

(middle) 1st pattern derived from PLS-regression analysis, represented as the regressions between SST anomalies and the first PLS component; C.I. = 0.1K

(lower) 2nd PLS pattern obtained from the prototype analysis

An iterative PLSR analysis can be performed as follows:

1) calculate the correlation coefficients between the first predictor field (say, SON SSTs) and the predictand (DJF precipitation);
2) obtain a time series (i.e., the first PLS component) by projecting the first predictor field onto the correlation map obtained in step 1;
3) use conventional least squares fitting and regress the predictand on the first PLS component to obtain the first partial regression;
4) linearly remove the first PLS component from both predictand and all predictor fields. The residual predictand and predictor fields become the new predictand and predictor fields;
5) repeat steps 1–4 to obtain higher ranked PLS components.

The final regression that links the predictor fields to the predictand is the *sum of all partial regressions* from the PLS components of each predictor.

Schedule and Milestones

Objectives	Tasks	Year 1	Year 2	Year 3
Advancement of ARLs		ARL2	ARL3	ARL4
Objective 1: Design, validate, and implement a prediction model for western US precipitation at S2S lead times	**Task 1:**	Consult partner organization, CDWR, on specific water management activity to be enhanced, solicit regular feedback, and demonstrate the functionality of the applied concept for potential use in decision-making		
	Task 2:	Extract sources of S2S predictability from PLS regression analysis	Develop and validate the statistical prediction model through assessment of model hindcast skill over an independent period	
	Task 3:	Apply ML methods on the extracted S2S sources of predictability	Develop ML-based predictive models after thorough cross-validation analyses and hyperparameter tuning for optimized performance	
Objective 2: Generate and communicate precipitation forecasts to the partner organization	**Task 4:**		Generate forecasts at the beginning of, and updated outlooks midway through the winter season to the partner organization; finish and provide documentation for the operation and maintenance of the forecast system	
Objective 3: Evaluate the representation of key sources of S2S predictability in operational prediction models	**Task 5:**		Compare and quantify dynamical model skill (from S2S Project and NMME models) relative to the skill from statistical and ML approaches; develop process-level insights into model strengths and shortcomings	

Thank you for listening!

Contact: agsengupta@ucsd.edu

Center for Western Weather and Water Extremes
SCRIPPS INSTITUTION OF OCEANOGRAPHY AT UC SAN DIEGO