Improving Statistical Prediction of Subseasonal CONUS Precipitation based on ENSO and the MJO by Training with Large Ensemble Climate Simulations

Cheng Zheng[1], Hyemi Kim[1,2], Emerson LaJoie[3], Shan He[1], Edmund K.M. Chang[1]

[1]*School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York, USA* [2]*Department of Science Education, Ewha Womans University, Seoul, Republic of Korea*

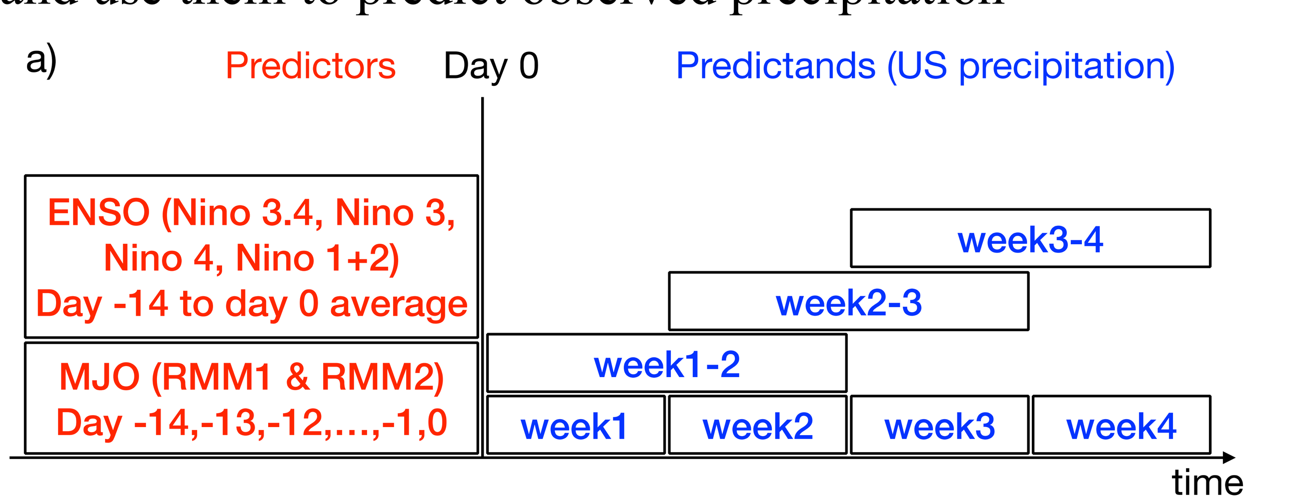
[3]*Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland, USA*

I. Background

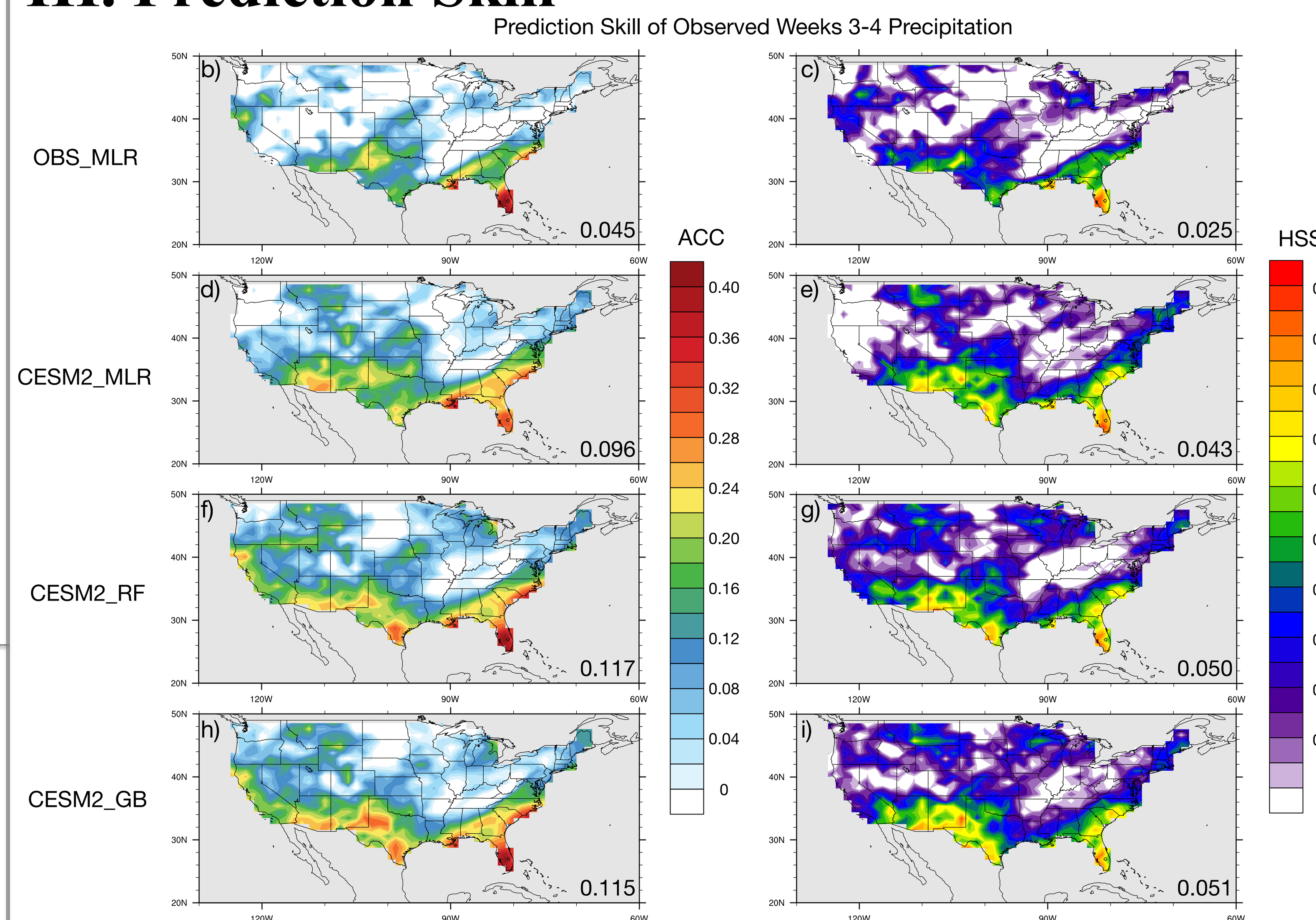
- El Niño-Southern Oscillation (ENSO) and the Madden-Julian Oscillation (MJO) are important source of predictability on S2S timescales.
- Statistical prediction tools using ENSO and the MJO can make skillful predictions (e.g., temperature) on subseasonal time scales.
- Scientific questions in this study: (1) By training statistical tools with large ensemble climate simulations, can we improve the prediction skill of grid point scale precipitation (compare to only training with observations)? (2) Can machine learning tools further improve the prediction skill?

II. Data and Methods

- Observations: Winter (DJF) CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS; RMM and ENSO indices (1982-2022)
- Large ensemble climate simulations: CESM2-LE, 100 ensemble members (1950-2022)
- Statistical/ML methods: multiple linear regression (MLR), random forests (RF), and gradient boosting (GB)
- Prediction skill metrics: anomaly correlation coefficient (ACC) and 3-category Heidke skill score (HSS)
- Baseline prediction skill: MLR tool trained with observational data using a leave-one-season-out cross validation (LOOCV) approach (OBS_MLR)
- Training/testing with CESM2-LE: Train prediction tools with 80 ensemble members (20 ensemble members reserved for testing). Freeze the tools (trained only with CESM2-LE) and use them to predict observed precipitation

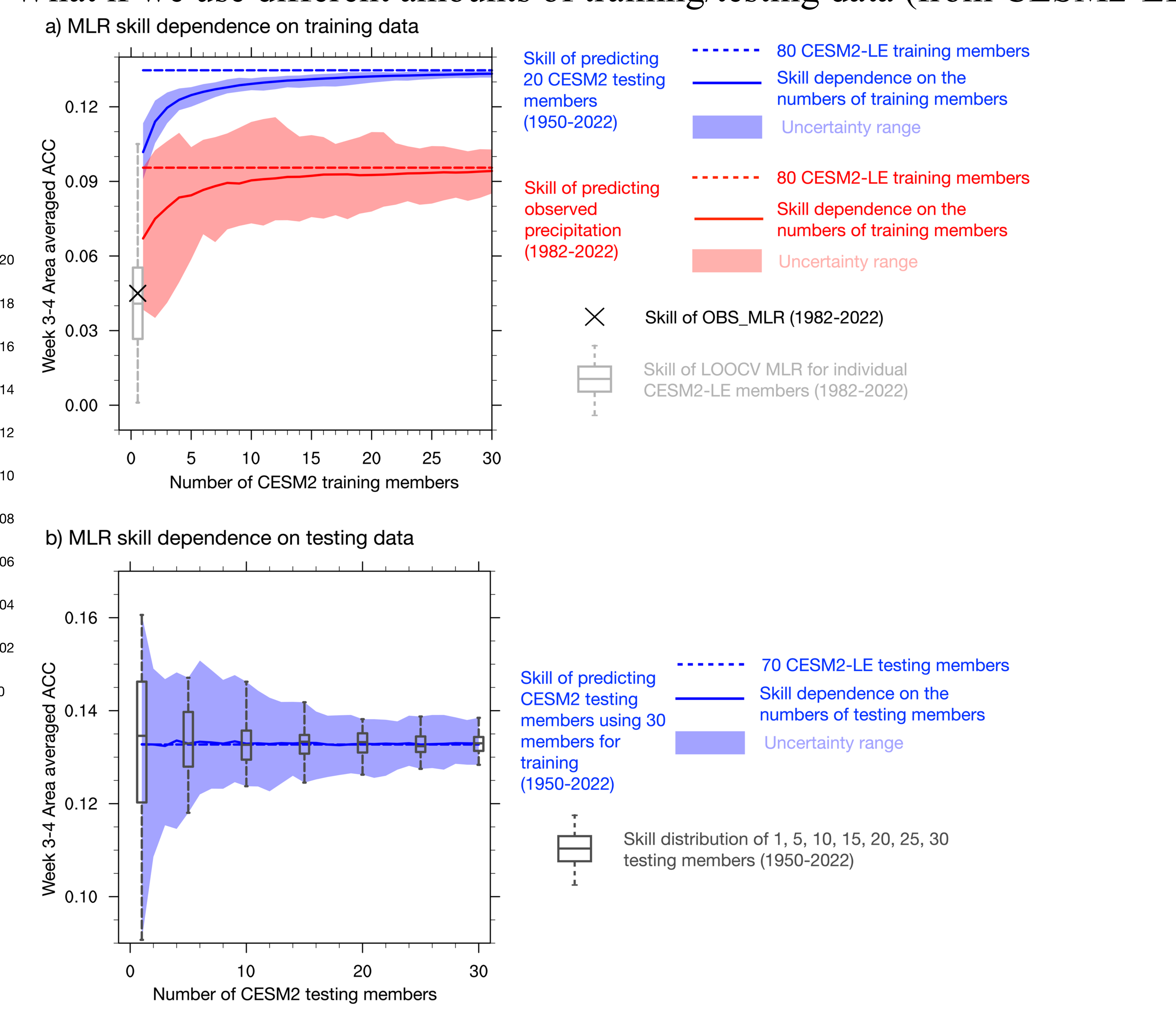


III. Prediction Skill

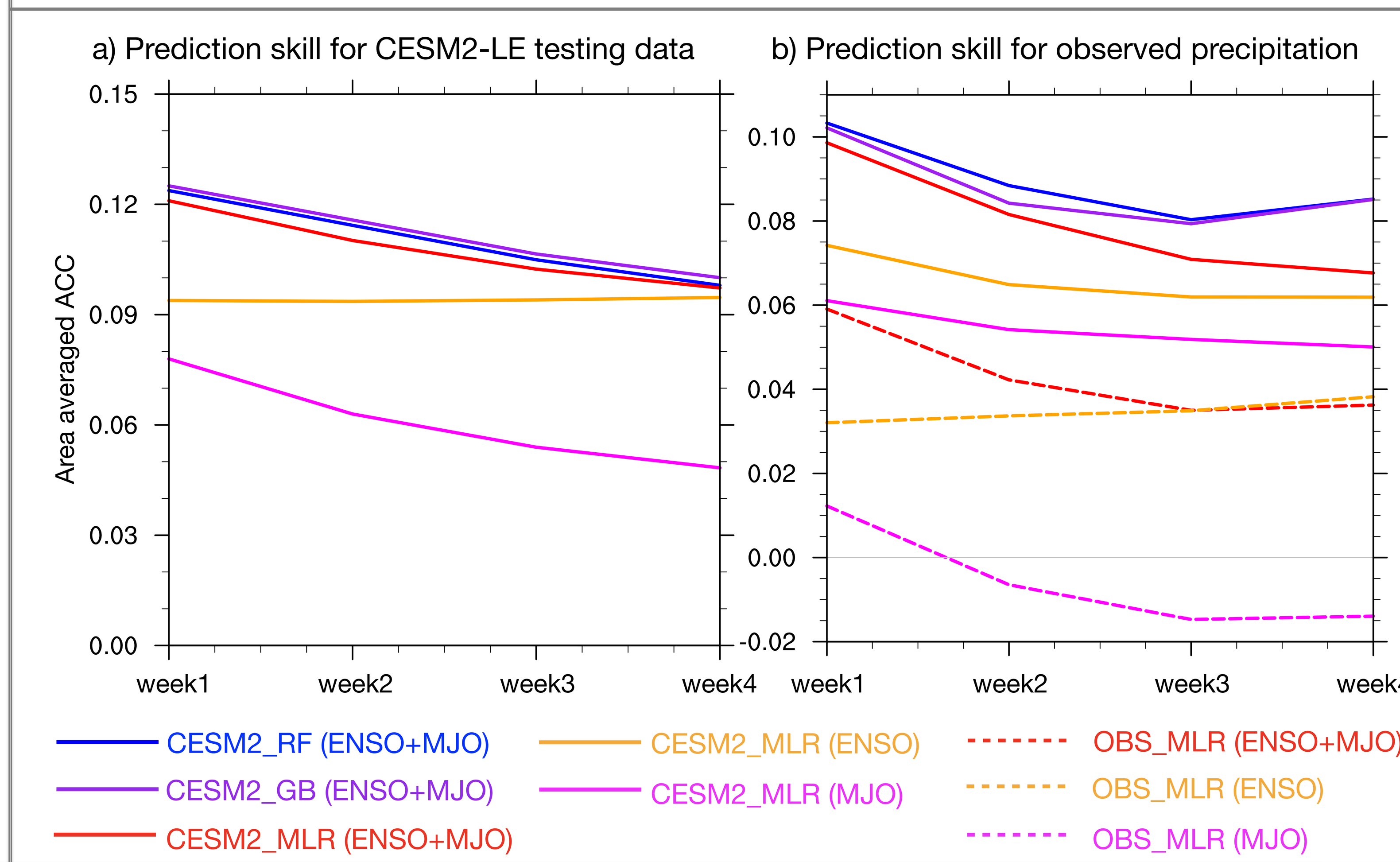


- Skill in predicting observed weeks 3-4 precipitation: OBS_MLR < CESM2_MLR < CESM2 machine learning tools
- By just using the linear method, there is a large gain in prediction skill by using CESM2 for training (CESM2_MLR) compared to OBS_MLR. Why?

What if we use different amounts of training/testing data (from CESM2-LE)?

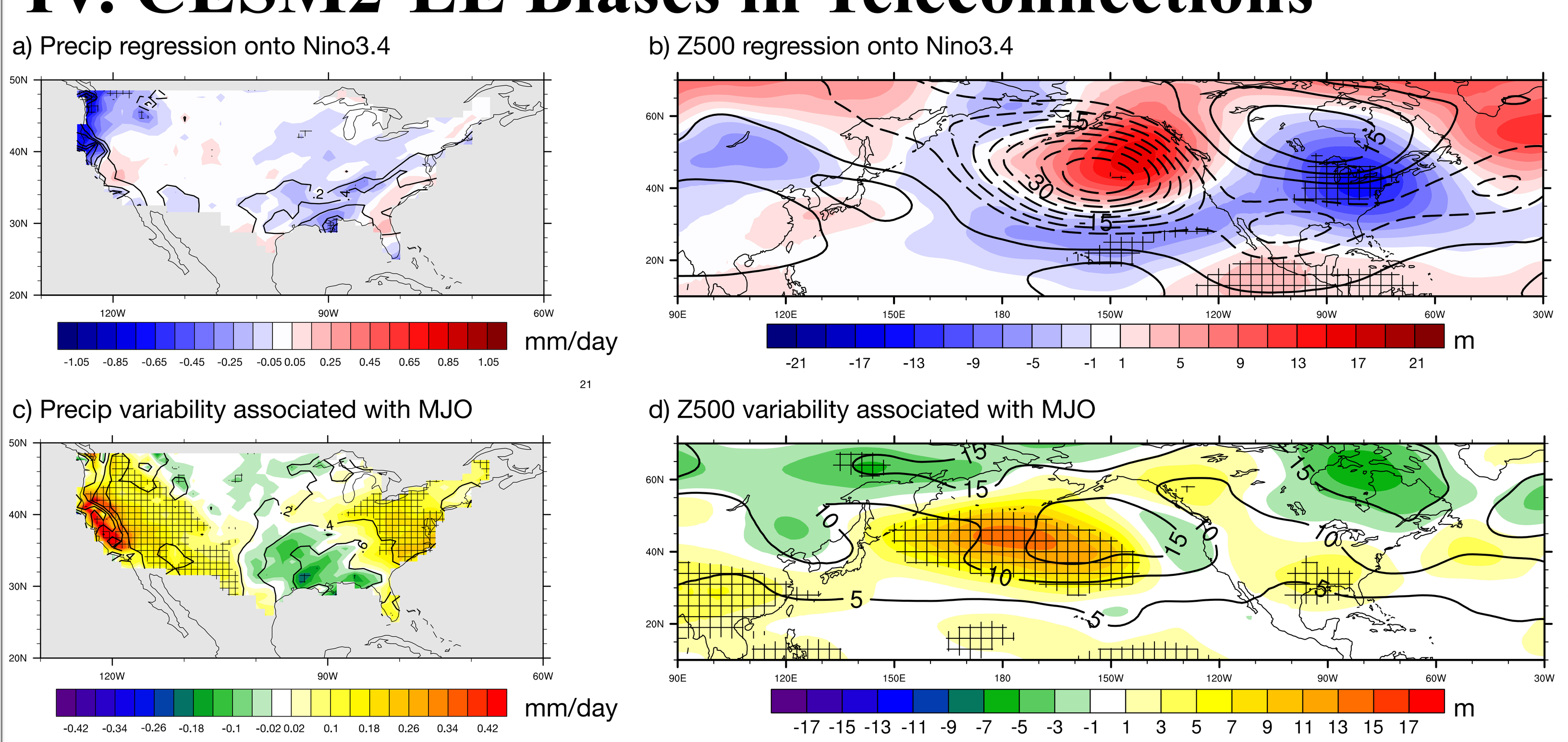


- Increasing the amount of training data (from CESM2-LE) improves the skill in predicting both observed and CESM2-LE precipitation (~2000 seasons of data required to establish robust statistical relationships).
  - Explains why OBS_MLR (LOOCV) skill is lower
- A substantial amount of validation data (~1500 seasons) is necessary to achieve a robust estimation of the prediction skill.
  - Large uncertainty in predicting skill due to limited availability of observational data



- ENSO has larger contribution to the prediction skill than the MJO
- The contribution from ENSO is relatively stable from week 1 to 4, while the contribution from the MJO decreases from week 1 to 4.
- Prediction tools trained with a large amount of data (CESM2-LE) always outperform those only trained with observational data.
- Machine learning tools can further improve the prediction skill compared to linear tools.

IV. CESM2-LE Biases in Teleconnections



Contours: Observed ENSO/MJO teleconnection (1982-2022)

Shadings: Differences in teleconnection between observation and CESM2-LE (1982-2022)

Crossed-hatched: Differences in teleconnection which cannot be solely explained by internal variability

- CESM2-LE has at least slight biases in ENSO teleconnection, and significant biases in the MJO teleconnection.
- Training with a large amount of biased climate simulation results in higher prediction skill than just training with observed data → A large amount of training data is important to reach high prediction skill

V. Conclusions

- Increasing the amount of training data improves the skill in predicting subseasonal precipitation, even if the training data is biased. About 2000 seasons of data are required to saturate the prediction skill.
- The scarcity of observational data may lead to significant uncertainties when evaluating prediction skills. Approximately 1500 seasons of validation data are necessary to achieve a robust estimation of the prediction skill.
- ENSO consistently contributes more to the prediction skill than the MJO. The contribution from ENSO remains stable from week 1 to 4, while the contribution from the MJO decreases with longer lead times.
- Machine learning tools can improve the prediction skill compared to linear methods.