

## Using data-driven methods to explore the predictability of surface soil moisture with FLUXNET site data

\*jinjing Pan<sup>1</sup>

1. Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-sen University.

Soil moisture (SM) is a key variable of land surface-atmosphere interactions. Data-driven methods have been widely used to predict SM, but the predictability of SM has not been well evaluated. This study investigated what variables and methods can be used to better predict SM for leading times of 7 days or longer with a global coverage of FLUXNET site data for the first time. Three machine-learning models, i.e., Bayesian linear regression, random forest and gradient boosting regression tree, are used for the prediction. Variables including atmospheric forcing, surface soil temperature, time variables (year, day of year (DOY) and hour), the Fourier transformation of time variables and lagged SM (7-14 day lagged) were sequentially added into models. A framework with five experiments is designed for factorial exploration of SM predictability. A stepwise method was used to build the best models for each site. The performance of regression models became better when adding more explaining variables in most cases. The results showed that from 47 % to 95% of variation of the best models can be explained. The important explaining variables are lagged surface SM, followed by DOY, year, soil temperature and atmospheric forcing. The predictability of SM depends highly on SM memory characteristics and the persistence of seasonality. The effect of SM memory characteristics on SM prediction as an initial condition question has been widely discussed in this paper. Our results also provide an insight that mechanisms of seasonality effects on SM should be also paid more attention to.

**Keywords:** A framework with five contrasting experiments is designed to investigate the factors affecting the predictability of soil moisture, The best models can explain between 47% and 95% of variation with different explaining variable combinations at different sites, The most important explaining variable is lagged surface soil moisture, followed by time variables, soil temperature and atmospheric forcing, The predictability of soil moisture depends highly on soil moisture memory characteristics and the persistence of seasonality