EGU2020-7428
https://doi.org/10.5194/egusphere-egu2020-7428
EGU General Assembly 2020

Augmenting satellite-derived soil moisture with multiple data streams using machine learning

Rene Orth and **Sungmin Oh**
Max Planck Institute for Biogeochemistry, Jena, Germany (rene.orth@bgc-jena.mpg.de)

Soil moisture plays a key role in land-atmosphere interactions through its influence on the energy and water cycles. Furthermore, its spatiotemporal variations can affect the development and persistence of extreme weather events. Consequently, soil moisture information is required for a wide range of research and applications, such as agricultural monitoring, flood and drought prediction, climate projection, and carbon-cycle modeling. Despite its scientific and societal importance, observations of soil moisture are sparse, in particular across time and at large spatial scales. Only models and satellite retrievals can provide global soil moisture information. While the ability of land surface models to represent the complex land-atmosphere interplay is still limited, satellite-based soil moisture data are a valuable alternative. However, these products suffer from a scaling based on models, and can only capture the top few centimeters of the soil.

In this study, we aim to augment satellite-based soil moisture data using machine learning. For this purpose we integrate satellite soil moisture with multiple hydro-meteorological data streams to derive global gridded soil moisture using Long Short-Term Memory (LSTM) neural networks. These networks are trained using in-situ soil moisture measurements as target data. With the resulting self-learned relationships, the LSTMs can produce in-situ-like soil moisture globally. We further analyze the implications of using point-scale target data to infer large scale information. The new dataset is derived separately for the surface and the deeper soil, thereby extending beyond the range covered by the satellite-based products. The integration of many data streams and multiple soil moisture observations through a powerful synergistic technique offers the potential to yield high accuracy. This is tested through rigorous cross-validation of the derived dataset. Finally, the planned datasets will permit consistent long-term, large-scale analysis to enhance our understanding of the hydrology-biosphere-climate interplay, to better constrain models and to support hydrological hazards monitoring and climate projections.