EGU24-12005, updated on 13 Jan 2025
https://doi.org/10.5194/egusphere-egu24-12005
EGU General Assembly 2024

A Machine Learning Framework for Extending SMOS Surface Soil Moisture Observations over Canada

Jeenu John, Laxmi Sushama, and Shinto Roose
1Department of Civil Engineering and Trottier Institute for Sustainability in Engineering and Design, McGill University

Continuous and long-term surface soil moisture (SSM) data is essential for advancing the understanding of land-atmospheric interactions and climate change studies. Despite the contributions of different satellite missions in acquiring SSM measurements, the presence of data gaps poses a significant challenge. In this study, a machine learning (ML) framework is developed to expand the Soil Moisture Ocean Salinity (SMOS) SSM observations in both spatial and temporal domains over Canada. In the initial phase of the proposed framework, ML models, including random forest (RF) and convolutional neural network (CNN), are trained and validated using SSM-relevant climatic and geophysical variables extracted from the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data (ERA5) for the 2011 to 2020 period and SMOS SSM for the same period as the target variable. While evaluating the developed models with unseen data from the years 2021 and 2022, the RF model shows slightly better performance when compared to that of CNN. The average root mean square error (RMSE) for RF is 0.0369 m^3/m^3 (Pearson correlation coefficient, R=0.94), while for CNN, the RMSE is 0.0494 m^3/m^3 (R= 0.89), with prediction biases mostly noted for regions with large inter-annual variability. Similarly, RF and CNN yield average RMSE values of 0.014 m^3/m^3 and 0.0635 m^3/m^3, respectively, when evaluated for spatial filling for the case of grid cells excluded during the training process. Hence, in the second phase of the proposed framework, the RF model is selected to extend the SMOS dataset for the 2008-2010 period. The temporal correlation analysis between the extended SMOS and Advanced Scatterometer (ASCAT) indicates reasonable correlations with values above 0.6, while the spatial correlation analysis reveals similar patterns between the two datasets, with smaller values for the summer season owing to the importance of local processes on SSM during this period. However, spatiotemporal extension of SSM to encompass surface types excluded during training remains a challenge and needs further studies. The developed framework holds the potential to address the spatio-temporal data gaps in other regions since both datasets are globally available.