Nicklas Simonsen

Soil moisture prediction with deep learning and remote sensing

This thesis develops a novel method for predicting soil moisture from remote sensing data using deep learning. Remote sensing of soil moisture creates better opportunity for applying the methods at a large spatial scale and deep learning creates the predictions by learning complex relationships from the data.

Soil moisture is a hydrological and climatic variable used to express the degree of saturation of the soil. The saturation of the soil affects various hydrological, climatic, and ecosystem functions and hence is an important variable in modelling, forecasting, and analysis within these fields. Soil moisture is traditionally measured using in situ probes, but this method only provides point-based measurements with poor spatial representation. Remote sensing is an increasingly popular alternative that can provide large-scale spatial estimations of soil moisture continuously. However, most remotely sensed solutions do not offer estimations below 1 km spatial resolutions, which is required for applications that focus on land covers with high heterogeneity, such as precision agriculture. Additionally, most of these solutions also do not provide complete time-series, which can be important for many applications. This thesis presents a new method of estimating soil moisture from remotely sensed products, such as Sentinel-1 microwave radar data and Sentinel-2 derived vegetation indices, in addition to climatic and topographic variables. This method relies on a recurrent neural network deep learning algorithm, that specializes in sequence data, where antecedent conditions are predictive of current and future conditions. The model is built as a semi-automatic pipeline that processes the various data types, trains the model, and creates an estimation. The entire pipeline and model are written in Python and uses the Tensorflow and Keras high-level libraries for the assembly of the model. Several variations of the dataset are tested to determine which variables are more beneficial to the model and if reducing the number of parameters in the model can increase the accuracy.

The results from the model show that static variables, which pertain variables that do not change over time, such as elevation and soil texture, are less useful to the model and result in lower accuracy when included. This is likely due to low number of stations used in the study, which are not numerous enough to represent a meaningful statistical distribution. The dynamic variables, which do change over time, such as soil moisture and temperature, are the most important in the model, as they contain more information due to their temporal nature. Additionally, the results show that some land covers have worse accuracy than others, hereunder forests due to the dense vegetation and peat bogs because of the high saturation of the soil. The results from the model are compared with existing soil moisture products, where it proves to be a competitive solution in accuracy, while also retaining a higher spatial and temporal resolution. This thesis concludes that deep learning is a promising but underexplored method for estimating soil moisture from remote sensing.

Keywords: Physical Geography and Ecosystem Analysis, Machine Learning, Deep Learning, Recurrent Neural Network, LSTM, Soil Moisture, Remote Sensing, SAR, Sentinel-1, Sentinel-2, ISMN, ICOS

Advisor: Zheng Duan
Master degree project 30 credits in Geomatics, 2021
Department of Physical Geography and Ecosystem Science, Lund University. Student thesis series INES nr xx.