EGU2020-342, updated on 04 Oct 2022
https://doi.org/10.5194/egusphere-egu2020-342
EGU General Assembly 2020

A Machine Learning Approach to Upscale Net Ecosystem Exchange to a Regional Scale: Integration of Eddy Covariance, Remote Sensing and Reanalysis Data

Oliver Reitz[1], Alexander Graf[2], Marius Schmidt[2], Gunnar Ketzler[1], and Michael Leuchner[1]
[1]Physical Geography and Climatology, RWTH Aachen University, Aachen, Germany (oliver.reitz@geo.rwth-aachen.de)
[2]Agrosphere Institute (IBG-3), Forschungszentrum Jülich, Jülich, Germany

Net Ecosystem Exchange (NEE) is an important factor regarding the impact of land use changes to the global carbon cycle and thus climate change. The Eddy Covariance technique is the most direct way of measuring CO_2 fluxes, however, it provides spatially discontinuous data from a sparse network of stations. Thus, generating high-resolution spatiotemporal products of carbon fluxes remains a major challenge. Machine Learning (ML) techniques are a promising approach to upscale this information to regional and global scales and can thereby help to produce better NEE datasets for earth-system modelling.

Our approach uses statistical relationships between NEE, vegetation indices and meteorological variables to train a Random Forest model with spatial feature selection to predict daily NEE values at 1 km spatial resolution for the Rur-catchment area (ca. 2400 km^2) in western Germany. Data from twelve Eddy stations of different land use types of the TERENO Network Eifel/Lower Rhine Valley between 2010 and 2018 were used to train and test the ML model. Factors potentially affecting NEE such as vegetation indices (NDVI, EVI, LAI) extracted from MODIS products, incoming solar radiation from Heliosat (SARAH-2) and additional meteorological variables from COSMO REA6 reanalysis products served as independent variables, which were further evaluated in regard to their relative importance for NEE prediction.

A novel spatial cross-validation scheme has been applied and compared to a conventional random k-fold cross-validation. This is important for the assessment of the model performance regarding spatial predictions beyond the scope of training locations in contrast to mere data reproduction. Results indicate a lower model performance evaluated with spatial cross-validation and that conventional random cross-validation hence leads to an overoptimistic view of the prediction skills. Nonetheless, the ML approach displayed a feasible way to upscale carbon fluxes to a regional scale utilizing different datasets and produced high-resolution NEE-raster for an entire catchment area.