EGU22-5624
https://doi.org/10.5194/egusphere-egu22-5624
EGU General Assembly 2022

Machine learning in four dimensions for mapping soil organic matter changes between 1953-2018 at 25m resolution in the Netherlands

Anatol Helfenstein[1,2], Vera Leatitia Mulder[1], Gerard B.M. Heuvelink[1], and Mirjam J.D. Hack-ten Broeke[2]

[1]Soil Geography and Landscape Group, Wageningen University, Wageningen, The Netherlands (anatol.helfenstein@wur.nl)
[2]Soil, Water and Land Use Team, Wageningen Environmental Research, Wageningen, The Netherlands

The European Green Deal emphasizes the importance of healthy soils for our planet and society. In order to monitor soil health, modelling soil organic matter (SOM) in space and over time is necessary to assess changes in the fertility of agricultural soils, combat climate change and maintain ecosystem services. In digital soil mapping, most recent statistical modelling approaches have used time series of remote sensing data, which became available from the early 1980s onwards, together with soil observations to make predictions in space and over time. While this has clear advantages, it does not provide spatially explicit, explanatory information for time periods before the 1980s, even though observations in soil databases are often from beforehand. In this study, we modelled changes in SOM in 3D space over 65 years on a national scale in the Netherlands. We used SOM observations from 345 000 locations from 0 to 2 m depth between 1953 and 2018. The covariates were comprised of proxies of soil forming factors either considered to be static (e.g. relief, parent material) or dynamic over these 65 years. As dynamic covariates, we used indices of digitized historic (1960 – 1980) and more recent (1986 – 2018) land use maps. These dynamic covariates were chosen for two reasons. Firstly, land use and land cover change are the main drivers of SOM change over the time period of several decades. This is especially true in the Netherlands, where the anthropogenic influence on soils has been tremendous. Approximately 82 % of the land surface are agricultural, urban or infrastructure areas, while 15 % consists of (managed) peatlands and up to 20 % has been reclaimed from the sea. Secondly, by including carefully mapped historic land use, we were able to take advantage of a longer time series of soil data to make space-time predictions of SOM over a longer time period. Predictions were made using the quantile regression forest (QRF) algorithm, whereby sampling depth and year were included during calibration. SOM predictions were validated in two ways: a) over the 65-year period using a 10-fold cross-validation and b) specifically for 1998 and 2018, where designed-based statistical inference was possible using a probability sample. We computed the mean error (ME), root mean squared error (RMSE), model efficiency coefficient (MEC) as accuracy metrics and the prediction interval coverage probability (PICP) as an evaluation of the prediction uncertainty. Results showed that spatial patterns were realistic and properly reproduced but that prediction of temporal dynamics was more challenging. This research is also of interest for spatio-

temporal soil modelling in other regions of the world that have soil data from the early and mid-19th century and historical land use and land cover data.