

EGU21-3368, updated on 19 Jun 2021

<https://doi.org/10.5194/egusphere-egu21-3368>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Spatial prediction of soil thickness with Gaussian Process Regression using pedological knowledge described by partial differential equations

Kerstin Rau^{1,3}, Thomas Gläzle^{3,2}, Tobias Rentschler^{1,4}, Philipp Hennig^{3,2}, and Thomas Scholten^{1,3,4}

¹Department of Geosciences, University of Tübingen, Tübingen, Germany

²Department of Computer Sciences, University of Tübingen, Tübingen, Germany

³Cluster of Excellence Machine Learning: New Perspectives for Science, University of Tübingen, Tübingen, Germany

⁴SFB 1070 ResourceCultures, University of Tübingen, Tübingen, Germany

Recently, there is a growing interest for the soil variable soil thickness in the soil science, geoscience and ecology communities. More and more scientists assume that soil thickness summarizes many different characteristics of the site that are important for plant growth, soil biodiversity and climate change. As such soil thickness can be an indicator of properties like water holding capacity, nutrient cycling, carbon storage, habitat for soil fauna and overall soil quality and productivity. At the same time, it takes a lot of effort to measure soil thickness, especially for larger and heterogeneous areas like mountain regions, which would require dense sampling. For these reasons, it is becoming increasingly important to spatially predict soil thickness as accurately as possible using models.

The typical difficulty in predicting variables in environmental sciences is the small number of samples in the field and resulting from this a small number of usable data points to train models in the spatial domain. One possibility to create valid models with sparse spatially distributed soil data is the combination of point measurements with domain knowledge. For soils and their properties such knowledge can be archived from related environmental data, for example, parent material and climate, and their spatial distribution neighboring the sample points. Frequently used machine learning methods for environmental modelling, especially in the geosciences, are the Gaussian Process Regression Models (GPRs), because a spatial correlation can already be implemented via the covariance kernel. One of the great advantages of using GPRs is the possibility to inform this algorithm directly with soil science knowledge. We can claim this knowledge in different ways.

In this paper we apply a new approach of implementing geographical knowledge into the Gaussian Processes by means of partial differential equations (PDEs), each describing a pedological process. These PDEs include information on how independent environmental variables influence the searched dependent variable. At first, we calculate for simple correlations between soil thickness and these variables, which we then convert into a PDE. As independent variables we initially use exclusively topographical variables derived from Digital Elevation Models (DEM) such as slope, different curvatures, aspect or the topographic wetness index. In this way, expert knowledge can

adapt the GPR model in addition to the already existing assumption of spatial dependency given by prior covariance, where near things are more related than distant.

The algorithm will be applied to a data set from Andalusia, Spain, developed by Tobias Rentschler. Among land use information gained from remote sensing, it contains our target variable soil thickness.