High-resolution spatial modelling of total soil depth for France

Marine Lacoste, Vera Leatitia Mulder, Nicolas Saby, Dominique Arrouays

High-resolution spatial modelling of total soil depth for France

M. Lacoste, V.L. Mulder, N. Saby, D. Arrouays

INRA, UR 0272 Science du sol, Centre de recherche d'Orléans, CS 40001 Ardon, 45075 Orléans cedex 2, France.
Tel: 0033 (0)2 38 41 80 32, email: marine.lacoste@orleans.inra.fr

Abstract:
Determining of total soil depth (SDT_t) is important because of its key role in supporting various ecosystem services and properties, including plant growth, water availability and carbon stocks. Therefore, the exhaustive mapping of total soil depth was set as requirement of the GlobalSoilMap project. However, determining SD_t proved to be hampered by the (1) high spatial variability, (2) estimation of the maximum soil depth of deep soils (> 1.5 m) and (3) discordance in the definition of SD_t.

The objective of this work was spatially explicit modelling of SD_t for France, following the directions of GlobalSoilMap, which requires modelling at 90m resolution. This was done by first modelling the general trend in soil depth, followed by kriging of the de-trended data. Considering the total surface area of France, being about 540K km^2, implies that the employed methods need to be able dealing with large data sets. This is especially challenging for geostatistical modelling. Furthermore, the general trend was assumed to be explained by the biotic and abiotic environmental conditions, as described by the Soil-Landscape paradigm. We adopted the definition of soil depth from the USDA Soil Survey Manuel, where soil depth is defined as the depth (in cm) to a lithic or paralithic contact.

The SD_t was determined for 2116 sites, originating from the French Soil Monitoring network (RMQS). This dataset encompasses a broad spectrum of climatic, soil and agricultural parameters and covers the entire metropolitan France, based on a regular, 16km x 16km grid. Soil depth ranged from 0 to 300 cm, with a mean value of 102 cm.
In addition, exhaustive environmental data was used for characterizing the climate (climate type, temperature and precipitations), the organisms (land use, forest type), the topography (Digital Elevation Model and its derived variables) and the other known soil properties (soil parent material, bare rock areas, soil waterlogging indices) – following the Soil-Landscape paradigm. Next, the general trend in the soil depth data was determined using the described data and boosted regression tree modeling (GBM). The model accuracy was assessed by an internal (R^2_{int} and $RMSE_{int}$) and 10-fold

cross-validation (R^2_{cross} and $RMSE_{cross}$). The residual term of the de-trended data was modelled, using multi-resolution Kriging based on Markov random fields. Typically, this method allows estimating the covariance models for large datasets, using a large number of basis functions, following a fixed rank Kriging approach. The method enables modelling spatial estimates for data having covariance models with small-nugget variances or strong spatial correlation over long ranges. Using this kriging approach, a spatially explicit uncertainty measurement was produced by computing the 90% prediction interval, as derived from the spatial estimates and their associated prediction uncertainty. Finally, the predicted SD_t was compared to an existing soil map (RRP, 1:250K) for a restricted area of France, as means of an external validation.

The GBM indicated that the general trend in SD_t was mainly explained by those variables derived from the DEM (40 % of relative importance, with elevation as the main variable); next, soil properties (26 % of relative importance, with soil parent material as the main variable); climate variables (23 % of relative importance, with mean maximum annual temperature as the main variable); and finally, land use data (11 % of relative importance, with forest type as the main variable). Model accuracies were found to be good for the internal validation (R^2_{int} = 0.72 and $RMSE_{int}$ = 32 cm) but were substantially lower for the cross-validation (R^2_{cross} = 0.20 and $RMSE_{cross}$ = 47 cm). As expected, the poorer results were obtained for higher values of SD_t, whereas the mean values were the best predicted. The additional assessment of the de-trended data, using the multi-resolution kriging, was found to be valuable for high-resolution modeling. Especially, the spatial explicit uncertainties support the interpretation of the model accuracy. Compared to previous SD_t models, this work proved to be better capable to capture the spatial variation and was found to be more consistent. Here, the multi-resolution kriging was key for modelling both small nugget variances and long-range spatial correlation. Improvements may be expected from using high-resolution biotic and abiotic environmental data, since it may improve modelling the soil-landscape interactions influencing soil pedogenesis. Concluding, this work provides a robust and reproducible method for high-resolution soil property modelling, in accordance with the GlobalSoilMap requirements.