The influence of covariates and sample density on digital soil mapping performance at a national scale

Azamat Suleymanov, Anne C Richer-De-Forges, Nicolas P. A. Saby, Dominique Arrouays, Manuel P Martin, Antonio Bispo

The influence of covariates and sample density on digital soil mapping performance at a national scale

Suleymanov Azamat[1*], Richer-de-Forges Anne C.[1], Saby Nicolas P.A.[1], Arrouays Dominique[1], Martin Manuel P.[1] and Bispo Antonio[1]

[1] *INRAE, Info&Sols, 45075 Orléans, France*

** Presenting author: azamat.suleymanov@inrae.fr*

In the context of Global Soil Nutrient and Nutrient Budget maps, the FAO Global Soil Partnership (GSP) initiated a country-driven digital soil mapping (DSM) approach. This involved predicting ten soil properties using national point data and a set of widely available covariates (GSP_Cov). In this study we demonstrated the impact of including additional national-based covariates and soil observations on prediction model performance, using mainland France as a pilot area. A Random Forest approach combined with the Boruta selection method was employed to map ten soil properties, including soil organic carbon, pH (water), total nitrogen, available phosphorus, available potassium, cation exchange capacity, bulk density, and texture (clay, silt, and sand). The GSP_Cov included common covariates representing terrain, climate, and organisms, whereas the second set included these covariates extended to additional national-level data such as existing soil and geological maps, remote sensing products, etc. Results showed notable enhancements in prediction performance for more than half of the properties, particularly for pH, CEC, and texture, whereas geological variables and previous pH maps significantly improved accuracy. Adding around 25,000 points to the learning dataset improved the performance of soil particle-size fraction predictions. This research emphasizes the importance of incorporating a diverse range of covariates at a national scale and densifying soil information to expand the feature and geographical spaces of multidimensional soil/covariates combinations.

Keywords: digital soil mapping, spatial covariates, sampling size, random forest