Geophysical Research Abstracts
Vol. 15, EGU2013-6899, 2013
EGU General Assembly 2013

Comparison of universal kriging and regression tree modelling for soil property mapping

Bas Kempen
Netherlands (bas.kempen@wur.nl)

Geostatistical modelling approaches have been dominating the field of digital soil mapping (DSM) since its inception in the early 1980s. In recent years, however, machine learning methods such as classification and regression trees, random forests, and neural networks have quickly gained popularity among researchers in the DSM community. The increased use of these methods has largely gone at the cost of geostatistical approaches.

Despite the apparent shift in the application of DSM methods from geostatistics to machine learning, quantitative comparisons of the prediction performance of these methods are largely lacking. The aims of this research, therefore, are: i) to map two soil properties (topsoil organic matter content and thickness of the peat layer in the soil profile) using regression tree (RT) modelling and universal kriging (UK), and ii) to compare the prediction performance of these methods with independent data obtained by probability sampling. Using such data for validation does not only yield a statistically valid and unbiased estimates of the map accuracy, but it also allows a statistical comparison of the accuracies of the maps generated by the two methods.

The topsoil organic matter content and the thickness of the peat layer were mapped for a 14,000 ha area in the province of Drenthe, The Netherlands. The calibration dataset contained soil property observations at 1,715 sites. The covariates used include layers derived from soil and paleogeography maps, land cover, relative elevation, drainage class, land reclamation period, elevation change, and historic land use. The validation dataset contained 125 observations selected by stratified simple random sampling of the study area. The root mean squared error (RMSE) of the soil organic matter map obtained by RT modelling was 0.603 log(%), that of the map obtained by UK 0.595 log(%). The difference in map accuracy was not significant ($p = 0.377$). The RMSE of the peat thickness map obtained by RT modelling was 24.9 cm, that of the map obtained by UK 21.7 cm. For this soil property, UK performed significantly better than RT ($p = 0.001$). Concluding, UK performed as good or better than RT modelling for mapping two key soil properties in a small area of the cultivated peatlands in The Netherlands. However, prediction performance of RT modelling might be further improved by using more advanced modelling methods such as boosted or bagged regression trees, or random forest models.