Mapping subsoil ripening using Bayesian Generalized Linear Modelling

Luc Steinbuch – Wageningen UR; ISRIC, Netherlands
Dick Brus – Wageningen UR, Netherlands
Gerard Heuvelink – Wageningen UR; ISRIC, Netherlands

One of the first soil forming processes in marine and fluviatile clay soils is ripening, the irreversible change of physical and chemical soil properties, especially consistency, under influence of air. We used Bayesian Binomial Logistic Regression (BBLR) to update the map of subsoil ripening for a reclamation area in the west of The Netherlands. Similar to conventional Binomial Logistic Regression (BLR), in BBLR the binary variable (the soil is ripened or unripened) is modelled by a Bernoulli distribution. The logit transform of the 'probability of success' parameter of the Bernoulli distribution is modelled as a linear combination of the covariates soil type, water ditch level and ground water table depth. To capture all available knowledge, Bayesian statistics combines two information streams, i.e. pre-observation knowledge summarized in a 'prior' probability distribution, and the actual observations. We combined subsets of different size (ranging from 5% to 50% of the original dataset of 676 observation locations) with different priors (based on legacy data) to investigate the effect of sample size and prior distribution of the quality of the map. The posterior parameter distributions, calculated by Markov Chain Monte Carlo simulation, vary in dispersion as well as in central values, especially for the smaller datasets. More informative priors decreased dispersion and pushed posterior central values towards prior central values. Interestingly, the resulting prediction maps were almost similar, probably because of the relative low predictive power of the BLR model. However, the associated prediction uncertainty maps were different: a more informative prior decreased prediction uncertainty. Validation showed that more informative priors resulted in stabilized and slightly better results, as quantified by the measure for separating power Area Under Curve; again this effect is more pronounced for smaller datasets. In our research, Bayesian Binomial Logistic Regression proved to be a flexible mapping tool, that – depending on the strength of the prior - stabilizes estimated parameters as well as prediction maps in case of small datasets.

keywords: *Bayesian Statistics, Informative priors, Clay soil ripening, Generalised Linear Model, Spatial prediction, Spatial prediction uncertainty*