Geophysical Research Abstracts,
Vol. 11, EGU2009-9708, 2009
EGU General Assembly 2009

A soil map of a large watershed in China: applying digital soil mapping in a data sparse region

F. Barthold (1), B. Blank (1), M. Wiesmeier (2), L. Breuer (1), and H.-G. Frede (1)

(1) Justus-Liebig-University Giessen, Institute of Landscape Ecology and Resources Management, Giessen, Germany (frauke.barthold@umwelt.uni-giessen.de), (2) Technical University Munich, Institute for Soil Science

Prediction of soil classes in data sparse regions is a major research challenge. With the advent of machine learning the possibilities to spatially predict soil classes have increased tremendously and given birth to new possibilities in soil mapping. Digital soil mapping is a research field that has been established during the last decades and has been accepted widely. We now need to develop tools to reduce the uncertainty in soil predictions. This is especially challenging in data sparse regions. One approach to do this is to implement soil taxonomic distance as a classification error criterion in classification and regression trees (CART) as suggested by Minasny et al. (Geoderma 142 (2007) 285-293). This approach assumes that the classification error should be larger between soils that are more dissimilar, i.e. differ in a larger number of soil properties, and smaller between more similar soils.

Our study area is the Xilin River Basin, which is located in central Inner Mongolia in China. It is characterized by semi arid climate conditions and is representative for the natural occurring steppe ecosystem. The study area comprises 3600 km2. We applied a random, stratified sampling design after McKenzie and Ryan (Geoderma 89 (1999) 67-94) with landuse and topography as stratifying variables. We defined 10 sampling classes, from each class 14 replicates were randomly drawn and sampled. The dataset was split into 100 soil profiles for training and 40 soil profiles for validation. We then applied classification and regression trees (CART) to quantify the relationships between soil classes and environmental covariates. The classification tree explained 75.5% of the variance with land use and geology as most important predictor variables. Among the 8 soil classes that we predicted, the Kastanozems cover most of the area. They are predominantly found in steppe areas. However, even some of the soils at sand dune sites, which were thought to show only little soil formation, can be classified as Kastanozems. Besides the Kastanozems, Regosols are most common at the sand dune sites as well as at sites that are defined as bare soil which are characterized by little or no vegetation. Gleysols are mostly found at sites in the vicinity of the Xilin river, which are connected to the groundwater. They can also be found in small valleys or depressions where subsurface waters from neighboring areas collect. The richest soils are found in mountain meadow areas. Pedogenetic conditions here are most favorable and lead to the formation of Chernozems with deep humic Ah horizons. Other soil types that occur in the study area are Arenosols, Calcisols, Cambisol and Phaeozems.

In addition, soil taxonomic distance is implemented into the decision tree procedure as a measure of classification error. The results of incorporating taxonomic distance as a loss function in the decision tree will be compared with the standard application of the decision tree.