Geophysical Research Abstracts
Vol. 16, EGU2014-10632, 2014
EGU General Assembly 2014

Integrating legacy soil information in a Digital Soil Mapping approach based on a modified conditioned Latin Hypercube Sampling design

Felix Stumpf, Karsten Schmidt, Thorsten Behrens, Sarah Schoenbrodt-Stitt, and Thomas Scholten
Tübingen, Germany (felix.stumpf@uni-tuebingen.de)

One crucial component of a Digital Soil Mapping (DSM) framework is outlined by geo-referenced soil observations. Nevertheless, highly informative legacy soil information, acquired by traditional soil surveys, is often neglected due to lacking accordance with specific statistical DSM designs. The focus of this study is to integrate legacy data into a state-of-the-art DSM approach, based on a modified conditioned Latin Hypercube Sampling (cLHS) design and Random Forest. Furthermore, by means of the cLHS modification the scope of actually unique cLHS sampling locations is widened in order to compensate limited accessability in the field. As well, the maximally stratified cLHS design is not diluted by the modification. Exemplarily the target variables of the modelling are represented by sand and clay fractions. The study site is a small mountainous hydrological catchment of 4.2 km^2 in the reservoir of the Three Gorges Dam in Central China.
The modification is accomplished by demarcating the histogram borders of each cLHS stratum, which are based on the multivariate cLHS feature space. Thereby, all potential sample locations per stratum are identified. This provides a possibility to integrate legacy data samples that match one of the newly created sample locations, and flexibility with respect to field accessibility. Consequently, six legacy data samples, taken from a total sample size of n = 30 were integrated into the sampling design and for all strata several potential sample locations are identified. The comparability of the modified and standard cLHS data sets is approved by (i) identifying their feature space coverage with respect to the cLHS stratifying variables, and (ii) by assessing the Random Forest accuracy estimates.