ELS2014 –The Earth Living Skin: Soil, Life and Climate Changes
EGU – SSS Conference
Bari | Italy | 22 – 25 September 2014
ELS2014-37-1
© Author(s) 2014. CC Attribution 3.0 License.

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

Felix Stumpf (1), Karsten Schmidt (1), Thorsten Behrens (1), Sarah Schoenbrodt-Stitt (1), Giovanni Buzzo (2), Christian Dumperth (3), and Thomas Scholten (1)

(1) Tübingen, Germany (felix.stumpf@uni-tuebingen.de), (2) University of Trier, Department of Environmental Remote Sensing and Geoinformatics, Chair of Spatial and Environmental Planning, Germany, (3) University of Erlangen-Nuremberg, Department of Geology and Mineralogy, Chair of Applied Geology, Germany

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

Felix Stumpf*[1], Karsten Schmidt[1], Thorsten Behrens[1], Sarah Schönbrodt- Stitt[1], Giovanni Buzzo[2], Christian Dumperth[3], Thomas Scholten[1]

[1]University of Tübingen, Department of Geosciences, Chair of Physical Geography and Soil Science, Germany
[2]University of Trier, Department of Remote Sensing, Chair of Spatial and Environmental Planning, Germany
[3]University of Erlangen-Nuremberg, Department of Geology and Mineralogy, Chair of Applied Geology, Germany
*Corresponding author: University of Tübingen, Department of Geoscience, Chair of Physical Geography and Soil Science, Rümelinstraße 19-23, Tübingen, Germany, Tel.: +49 7071 29 73942, e-mail: felix.stumpf@uni-tuebingen.de

Abstract
Within the framework of the joint Sino-German project YANGTZE GEO, the subproject “Soil Erosion” aims to identify soil erosion risks and sediment transport pathways into the reservoir of the Three Gorges Dam in Central China. For this purpose quantitative soil information is of crucial importance.
The study focuses on setting up a process-oriented soil erosion model by developing a method to obtain soil property data efficiently and making use of available soil legacy data.
The training area is a catchment of 4.2 km^2, about 70 km upstream the dam, that is assumed to adequately represent the reservoir region.
Digital Soil Mapping (DSM) represents a bunch of methods to estimate spatially distributed soil property information by combining field-obtained soil information with area covering predictor covariables through equations or classification rules. In this context, the sampling design for the field-obtained soil observations outlines a crucial control factor. Additionally, highly informative legacy soil information is often neglected due to lacking accordance with specific statistical and target variable oriented DSM sampling designs. Hence, we increased the efficiency of a state-of-the-art Random Forest (RF) approach by integrating legacy soil data through a modified conditioned Latin Hypercube Sampling (cLHS) design. Furthermore, by means of the cLHS modification we widened the scope of actually unique cLHS locations in order to compensate limited accessability in the terrain with respect to field sampling . Exemplarily the target variables of the Random Forest modelling are represented by the top soil sand fractions coarse sand (2 mm – 0.63 mm), medium sand (0.63 mm – 0.2 mm) and fine sand (0.2 mm – 0.063 mm). The cLHS modification is accomplished by demarcating the histogram borders of each 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. To assess the quality of the approach the RF processing is conducted with three different field-obtained input data sets; the modified cLHS data (i) without and (ii) with integrated legacy data, and (iii) the pure legacy data The approach is comparatively validated by identifying the feature space coverage of the input data sets with respect to the cLHS stratifying variables, and by comparing the RF accuracy estimates. Additionally, the accuracy estimates of both cLHS and legacy data models with a samples size n = 30 are compared to the model, processed with all available soil observations (n = 65).
The modification enables to substitute 20 % of cLHS samples by legacy data and a variable number of alternative

cLHS locations per stratum were identified. For all target variables, except fine sand, the cLHS models outline similar results, while outcompeting the legacy data models. Moreover, the cLHS models, except for fine sand, are similar to the models, based on all observations.