Geophysical Research Abstracts
Vol. 17, EGU2015-2384, 2015
EGU General Assembly 2015

Assessing efficiency of spatial sampling using combined coverage analysis in geographical and feature spaces

Tomislav Hengl
ISRIC — World Soil Information, Wageningen, Netherlands (tom.hengl@wur.nl)

Efficiency of spatial sampling largely determines success of model building. This is especially important for geostatistical mapping where an initial sampling plan should provide a good representation or coverage of both geographical (defined by the study area mask map) and feature space (defined by the multi-dimensional covariates). Otherwise the model will need to extrapolate and, hence, the overall uncertainty of the predictions will be high. In many cases, geostatisticians use point data sets which are produced using unknown or inconsistent sampling algorithms. Many point data sets in environmental sciences suffer from spatial clustering and systematic omission of feature space. But how to quantify these "representation" problems and how to incorporate this knowledge into model building? The author has developed a generic function called "spsample.prob" (Global Soil Information Facilities package for R) and which simultaneously determines (effective) inclusion probabilities as an average between the kernel density estimation (geographical spreading of points; analysed using the spatstat package in R) and MaxEnt analysis (feature space spreading of points; analysed using the MaxEnt software used primarily for species distribution modelling). The output 'iprob' map indicates whether the sampling plan has systematically missed some important locations and/or features, and can also be used as an input for geostatistical modelling e.g. as a weight map for geostatistical model fitting. The spsample.prob function can also be used in combination with the accessibility analysis (cost of field survey are usually function of distance from the road network, slope and land cover) to allow for simultaneous maximization of average inclusion probabilities and minimization of total survey costs. The author postulates that, by estimating effective inclusion probabilities using combined geographical and feature space analysis, and by comparing survey costs to representation efficiency, an optimal initial sampling plan can be produced which satisfies both criteria: (a) good representation (i.e. within a tolerance threshold), and (b) minimized survey costs. This sampling analysis framework could become especially interesting for generating sampling plans in new areas e.g. for which no previous spatial prediction model exists. The presentation includes data processing demos with standard soil sampling data sets Ebergotzen (Germany) and Edgeroi (Australia), also available via the GSIF package.