Geophysical Research Abstracts
Vol. 17, EGU2015-13562, 2015
EGU General Assembly 2015

Prediction interval evaluation in modelling of soil texture for regional mapping: methodology and a case study

Rossano Ciampalini (1,2), Manuel Martin (2), Nicolas P.A. Saby (2), Anne C. Richer de Forges (2), Pierre Nehlig (3), Guillaume Martelet (3), and Dominique Arrouays (2)

(1) Cardiff University, School of earth and ocean sciences, Cardiff, United Kingdom (ciampalinir@cardiff.ac.uk), (2) INRA, US InfoSol - Centre de recherche d'Orléans, ORLEANS, France, (3) BRGM, Orléans, France

Model uncertainty mapping represents a practical way to describe efficiency and limits of models prediction and can be calculated using different techniques. The object of this study is to determine and apply a procedure for the prediction interval (PI) evaluation for extended maps of soil granulometric fractions (i.e. clay, silt, sand) in the region "Centre" of France.

Among the various methodologies for PI determination, a recent approach is the use of a non-parametric procedure evaluating the prediction interval. The PI is defined as a conventional bound of the predicted values (i.e. 95th percentile) and can be calculated as follows. Assuming a relationship between the inputs of the model and the resulting prediction error (Shrestha et al., 2006, Malone et al., 2011), the input variables-space is classified into different clusters having similar errors with a fuzzy c-means clustering technique. Then, a prediction interval (PI) is calculated for each cluster on the basis of the associated empirical distributions of the errors and considering the degree of membership belonging to each cluster. A relationship between the input variables and the computed prediction intervals is founded using a modelling procedure (calibration), then; the relationship is applied to estimate the prediction interval for the out-of-sample data (validation) (e.g. Solomatine et al., 2008, 2009, Malone et al., 2011). This approach requires the assumption of a relationships between the input variables and the errors, and, obviously the relevancy and accuracy of such approach depends on the validity of the assumption. These assumptions have been accepted in all the studies quoted above.
In this work we adopted a similar procedure to the third approach. Our hypothesis is, if a correspondence is supposed and identified between confidence interval and predictors (i.e. 2.5-97.5% values, respectively), a model between predictors and PI may be used to extrapolate it to the whole map. This approach is similar the Malone et al., (2011) here the correspondence is not just hypothesized but is validated within the modelling procedure we adopted with an internal leave out cross validation procedure as follows: 1) a leave-out cross validation iterative procedure (i.e. 90-10%) is adopted to build an empirical distribution of the model prediction on each validation point, 2) at each data point, the prediction interval is evaluated for the required intervals (i.e. 2.5-97.5% of the predicted values), 3) BRT models between lower or upper PI and covariates is built and a procedure of cross-validation for the evaluation of the goodness of fit is performed. After all, the resulting models are applied to the whole map dataset to extrapolate the prediction intervals.
Comparisons between wideness and type of the distribution of the PI is done comparing two confidence levels 0.90 and 0.95 as the most currently used bounds.