Sampling design method details

Sampling algorithm: Four algorithms are available. The directed and stratified directed algorithm uses the Surface Reconstructor algorithm (see details below) and requires a raster dataset. Regular grid sampling and random stratified sampling are also available. For these only an area to sample needs to be delineated (Area of interest tab).

The Surface Reconstructor algorithm for directed sampling uses a raster dataset to find optimal sample locations. The sampling strategy is based on the principle that an interpolation of the samples should be as similar as possible to the guide raster. When sample locations are identified, first the center point of the raster cell with the maximum deviation from the covariate raster mean is sampled. Then the raster cell with the maximum deviation from the first sampled raster cell is sampled. From then on, the values of the sampled raster cells are interpolated by inverse distance weighting (idw power = 2; no of neighbours = 8) and the center point of the raster cell with the largest absolute difference to the guide raster (error) is sampled. A new idw interpolation is made and a new cell is sampled. This is repeated until the specified number of samples is reached or one sample is placed in each cell of the stratification grid.

The method was first described by Olsson (2002) and is also implemented in the R package SurfaceTortoise.

Grid size: A positive number. Cell side (metres) of a square stratification grid. If not specified, and a stratified sampling design is chosen, a grid size will be computed from the specified no of samples and the sampled area. Note that only one of the inputs grid size or no. of samples can be specified at a time. For the stratified directed and the random stratified sampling, the likelihood for a clipped stratum, e.g. at the edge of the area to be sampled, is equal to the area of that stratum divided by the area of a full stratum. For the regular grid sampling, a clipped stratum is sampled if the center point of the unclipped sample is covered. Valid for all stratified sampling algorithms (i.e. all sampling algorithms but the directed sampling).

No. of samples: An integer. The number of samples to place. If not provided, the no of samples to place will be computed from the Grid size and the size of the area. Note that only one of the inputs grid size or no. of samples can be specified at a time. Valid for all sampling algorithms.

Min distance: A positive number. Minimum distance (metres) allowed between samples. Valid for the 'random' and the 'directed' methods (i.e. all sampling algorithms but the regular grid sampling).

Buffer along edge: A positive number. Buffer zone (metres) inside the sampled area border, where sampling is prohibited. Optional for all sampling designs.

Local map adaptation method details

This function is intended for local adaptation and evaluation of large extent digital soil maps. A raster map and an area of interest are required These are specified on the area of interest tab. In addition point location data must be uploaded. The values in the attribute column must be of the same entity and have the same unit as the chosen raster layer. Uploaded shapefuls must have the wgs1984 reference system and there must be columns with longitude and latitude values (decimal degrees), both in shapefiles and textfiles. However, as the analyses needs to be carried in a projected coordinate system, all spatial datasets will be projected onto the Web Mercator coordinate system (epsg: 3857) before the analyses. T avaoid excessive extrapolation, the specified area of interest will be clipped the buffered point locations. The buffer width is 1.5 times the largest distance between one point and its nearest neighbour.

Four maps are (created and) evaluated: 1) the original raster map, 2) a map created solely based on the soil samples data (ordinary kriging using a standardized variogram), and 3-4) two maps based on a combination of the raster data and the point observations (regression kriging and residual kriging, both using standardized variograms). The original raster map is evaluated by comparing the raster values with the uploaded point observations (independent validation), while the new maps are evaluated by leave-one-out cross validation. A number of evaluation measures are computed: the Nash-Sutcliffe modelling efficiency (E; Nash & Sutcliffe, 1970), the mean absolute error (MAE; Janssen & Heuberger, 1995), the coefficient of determination of a linear regression between predicted and measured values (r2) and the bias (ME).

Specifications of the standardized variograms models:

Model:	Gaussian
Range:	Half of the square root of the mapped area
Sill:	Variance of observed values
Nugget:	0.1 * sill

The mapsRInteractive algorithmns have been described ad by Piikki et al. (2017) and Nijbroek et al. (2018). More details can be found in these publications. It is also implemented in the R package mapsRinteractive.

References

Olsson, D. 2002. A method to optimize soil sampling from ancillary data. Poster presenterad at: NJF seminar no. 336, Implementation of Precision Farming in Practical Agriculture, 10-12 June 2002, Skara, Sweden.

Nijbroek, R., Piikki, K., Söderström, M., Kempen, B., Turner, K. G., Hengari, S., & Mutua, J. (2018). Soil Organic Carbon Baselines for Land Degradation Neutrality: Map Accuracy and Cost Tradeoffs with Respect to Complexity in Otjozondjupa, Namibia. Sustainability, 10(5), 1610. doi:10.3390/su10051610.

Piikki, K.,Söderström, M., Stadig, H. 2017. Local adaptation of a national digital soil map for use in precision agriculture. Adv. Anim. Biosci. 8, 430–432.

Janssen, P.H.M.; Heuberger, P.S.C. 1995. Calibration of process-oriented models. Ecol. Model., 831, 55–66.

Nash, J.E.; Sutcliffe, J.V. 1970. River flow forecasting through conceptual models. part I—A discussion of principles. J. Hydrol., 103, 282–290.