Spatial Coverage Sampling and Random Sampling from Compact Geographical Strata in R

Dennis Walvoort, Dick Brus & Jaap de Gruijter

Introduction

Both for mapping and for estimating statistics like the spatial mean of environmental variables, the precision will usually be increased by dispersing the sampling locations so that they cover the study area as uniformly as possible. This can be achieved by means of the `spcosa`-package, a new add-on package for R (R Development Core Team, 2009).

Algorithms

Sampling locations are distributed evenly over the study area by selecting them in compact spatial strata. Compact strata can be constructed by applying k-means clustering to the coordinates of a fine grid representing the study area. Two k-means algorithms have been implemented in the `spcosa`-package: a transfer and a swopping algorithm. Both result in compact strata. However, the latter also keeps the areas identical.

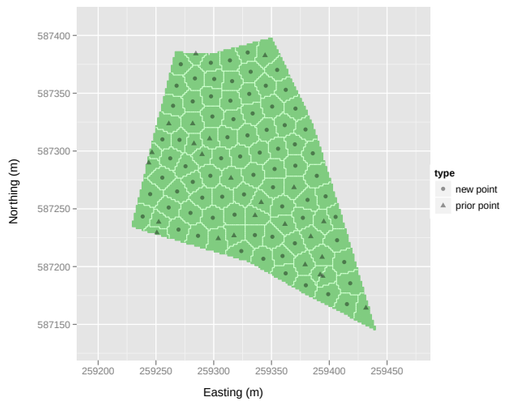
Examples

Spatial coverage sampling *without* prior points

In the figure below, the `spcosa`-package has been used to evenly distribute 100 sampling locations over the study area. This sampling pattern is useful for mapping.

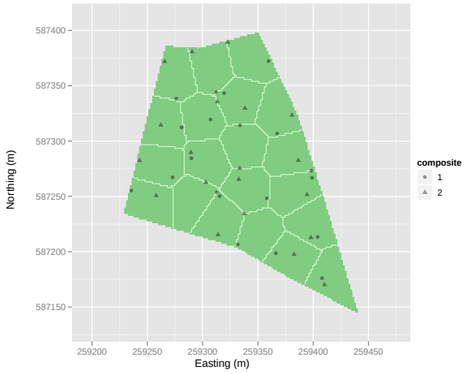
Spatial coverage sampling *with* prior points

In the figure below, the `spcosa`-package has been used to add 75 new locations (dots) to an existing configuration of 25 locations (triangles). New locations have been evenly distributed over the area while taking existing locations into account.



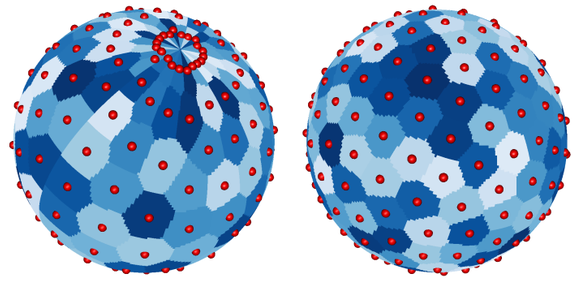
Random sampling for composites

In the figure below, the `spcosa`-package has been used to select two composites by stratified random sampling for design-based estimation of the field mean. The strata are forced to be of equal area to facilitate fieldwork. See Brus *et al.* (1999) and de Gruijter *et al.* (2006) for more details.



Spatial coverage sampling on a sphere

At spatial scales where the curvature of the Earth's surface cannot be ignored, the `spcosa`-package uses great circle distances (right figure) in stead of Euclidean distances (left figure). As evidenced by the left figure, Euclidean distances are clearly not appropriate at these scales due to pronounced edge effects near the poles and at 180° longitude (Walvoort *et al.*, 2009).



Availability

`spcosa` is an add–on package of R (R Development Core Team, 2009). It is available at the Comprehensive R Archive Network (`http://cran.r-project.org`).

References

Brus, D. J., Spätjens, L. E. E. M., and de Gruijter, J. J. (1999). A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. Geoderma 89: 129-148

de Gruijter, J., D. Brus, M. Bierkens, M. Knotters (2006). Sampling for Natural Resource Monitoring. Springer, Berlin, 332 pp.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org`.

Walvoort, D., D. Brus and J. de Gruijter (2009). Spatial Coverage Sampling on Various Spatial Scales. Pedometron 26: 20–22

*corresponding author (✉ dennis.walvoort@wur.nl)