266

An adaptive geostatistical sampling design for biodiversity studies via citizen science

Biodiversity estimates using data collected in citizen science studies typically showcase opportunistic sampling bias, often partly caused by a mechanism termed preferential sampling. This represents processes that generate sampling locations that are stochastically dependent on the spatial process that underlies the biodiversity dynamics. Preferential sampling can be corrected for by fitting a joint generalized linear mixed model where a geostatistical model of the outcome of interest shares a spatial random effect with a log-Gaussian Cox process that models the choice of sampling locations. However, data collection in citizen science is often simultaneously driven by multiple latent sampling processes that are difficult to disentangle, which often renders the model-based corrections still suboptimal.

We therefore could attempt to improve data quality at the sampling stage. A lattice-plus-close-pairs (LPCP) design is a geostatistical randomisation method consisting of a randomised systematic grid design, which allows for the optimal estimation of the spatial parameters in geostatistical models, in combination with a set of sampling points that lie in close proximity to each other, to efficiently estimate small-scale spatial variation. Due to its labour intensiveness, this sampling design is not recommended in a citizen science context. We have therefore developed an adaptive spatio-temporal LPCP design, which is set up as follows: (i) at time 0, a spatial LPCP design is constructed and citizens are asked to collect data at a set of locations; (ii) based on the adherence to the protocol, a preferential or non-preferential geostatistical model is fitted to the data and prediction uncertainty is evaluated; (iii) at time 1, the randomisation process is updated such that data points on locations with larger prediction uncertainties obtain more sampling priority. Afterwards, (ii) and (iii) are updated iteratively until the end of the study period. This design is evaluated using simulation studies, based on Belgian species richness data collected by LIKONA, a local citizen science organization. It is implemented in a LIKONA study on invertebrate monitoring. Our results show considerable improvement of species richness estimates when the adaptive LCPC design is implemented in comparison to those coming from data that were collected without adhering to a protocol.