Adaptive sampling for community science: can both models and people benefit?

Community or citizen science represents a major source of biodiversity information and many researchers are currently working with communities to collect data in the most informative locations. Increasingly, this effort is model-based rather than empirical i.e. models of existing information are used to inform the best new locations for participants to visit through an adaptive sampling design. If the aim is to produce spatial estimates of species occurrence then, theoretically, selecting new locations based on prediction uncertainty of an initial model should lead to reductions in average predictive variance in an updated model compared to random sampling. However, applying adaptive sampling in the context of citizen science raises complications which may alter the potential benefit from adaptive designs. For example, initial data may be a biased sample of geographic or environmental space, and not all space may be accessible to the community. The efficacy of adaptive sampling may also be reduced as recorders can decide whether or not to follow recommendations from an adaptive design. Finally, we consider the likelihood of a tradeoff between the objectives of model improvement versus the need to maintain recorder engagement.

We explore the potential benefits of adaptive sampling in the context of citizen science in a simulation study parameterised to reflect current recording patterns of British butterflies. We consider six potential adaptive designs ranging from empirical (record where no records currently exist) to model-based (record where uncertainty is highest). Some of the designs also consider incentives to recording (record where rare species are most likely to be seen). We find that adaptive sampling in the realistic context of biased initial data and variable uptake is rarely a clear improvement over existing sampling, and tradeoffs may occur which could risk both the accuracy of model outputs and future engagement of recorders. These findings are important when considering how to optimise recording effort in the future.