Strategies for Bayesian Optimal Adaptive Design for the Collection of Citizen Science Data

Max Savery[1] and Stijn Luca[1]

[1]Ghent University, Department of Data Analysis and Mathematical Modelling, Ghent, Belgium

Citizen science data has proved to be a valuable supplement to traditional planned survey data, which is generally labour-intensive and costly to collect. In parallel with the uptake of this data source there has been a surge in methodology that can accommodate such data, particularly in order to account for the bias that is inherent to its opportunistic collection. However, there has been notably less focus on methods for the design of surveys that can recommend potential locations for citizen scientists to visit. In this work we take advantage of the comprehensive framework of Bayesian optimal adaptive design and make refinements necessary for its use in the design of adaptive surveys for citizen scientists. Due to the computational complexity of Bayesian design, there are a number of approximations that can be used in practice, many of which may be necessarily bespoke to the application of interest. In order to explore the effect of such approximations for citizen science applications, we compare multiple adaptive design strategies, including both simpler model- or design-based criteria and those that are more fully Bayesian. Furthermore, we propose a novel strategy for the adaptive design of citizen science surveys that assesses the information contribution of each proposed location, in order to find a compromise between the speed of simpler evaluative metrics and the robustness of the computationally intensive Bayesian design methods. We optimize the designs based on a flexible utility function that considers model-based criteria such as the Kullback-Leibler divergence and citizen scientist preference such as accessibility of the locations. Importantly, we also perform a sensitivity analysis on the accuracy of the approximation to the expectation of the utility function and its effect on the obtained designs. In order to illustrate the effect of the different strategies on the designs and the modelling outcomes, we work in a virtual environment with species observations simulated from real-world covariate data. By putting these aspects of optimal design and citizen science data collection together in a single work, we seek to make the methods of Bayesian adaptive design more transparent and accessible to both citizen scientists and organizations that plan citizen science projects. As this project is currently in progress, we hope to elicit feedback regarding potential use-cases for our approach, as well as situations where adaptive design may fail to be practical.