

A probabilistic generating process of Citizen Science data: Modelling and Estimation of Parameters.

Citizen Science refers to the open engagement of the public in scientific activities. For example, several biodiversity projects encourage regular citizens to report the species they observe. Enabling citizens to feed databases enlarges the spatial coverage and the temporal resolution of biodiversity data. However, it comes at the risk of having biased unstructured sampling designs, information focused on more easily detectable species and misspecification of the species observed, among others.

Given the inherent biases to Citizen Science data, making inferences with them needs to account for these biases. In order to do it, we propose a probabilistic generating process of Citizen Science data. It starts by assuming that the true occurrence of species is a Spatial Point Pattern. Then, three typical sources of bias in Citizen Science are regarded as sources of thinning for the true point pattern. These are the sampling process of Citizen Scientists; the detection probability which is characteristic of each species, and the misclassification of the observed species.

Our proposed model is general and flexible, providing Citizen Science data analysts with the tools to account for any source of bias in their data. This is made possible by using a Bayesian approach as we aim to make inferences about key parameters that explain the ecological process of each species. We simulate a three-stage thinning procedure for the occurrences of a given species and we will make use of the model we propose to estimate the true ecological parameters as we account for these sources of biases.