

Understand Observer's behavior in Citizen science data collection to better correct for sampling bias

In recent years, the increase of data availability with citizen science campaigns have raised questions on its quality. An observer's record is part of a sample of potential observations from a species true distribution in space. Species distribution models (SDMs) can be severely impacted by non-random spatial distribution of records. Multiple methods exist to correct for spatial bias such as background modification, modelling of the bias or data modification. Most methods imply that the sampling is uneven in space and determined by the observers' choices of where to search for observations. Road networks are often presented as the cause of presence only data bias such that observers do not deviate far from road networks in making observations. One common correction method is to calculate distances to the road network as a proxy for the bias and correcting for this bias by setting the modelled covariate equal to a common value (Warton et al., 2013). However, this correction implies that each observer behaves in the same manner, which in practice may not be the case. We can differentiate two common observer behaviors: exploring and following. Under this paradigm, explorers do not always follow the road network and will seek to observe species in new places far away from other observations. By contrast, followers will target observations close to already observed species locations and will stay closer to the road network. As such, it is worth investigating whether the Warton et al approach to correcting for observer bias holds under varying observer behaviors, or whether a data-driven approach based on modelled observer behavior may lead to better predictions. To do so, we used the new software platform obsimulator developed by Dr Panu Somervuo (University of Helsinki) to simulate patterns of points driven by observer behavior. We developed two correction methods based on the Warton et al. approach, that use k-nearest neighbours and density calculation to evaluate the sampling bias correction for different observer behavior.