Accounting for biogeographical ignorance within biodiversity modelling

Biodiversity data can be analysed to predict species distribution at various scales of time and space. However, survey completeness and temporal decay in data quality introduce uncertainty into biodiversity models. Researchers Joaquín Hortal, Juliana Stropp (National Museum of Natural Sciences, Spain), Richard Ladle (University of Porto, Portugal), and Geiziane Tessarolo (State University of Goiás, Brazil), among others, are constructing the first Maps of Biogeographical Ignorance (MoBi) that account for uncertainty in biodiversity analysis. Presented alongside species distribution models (SDMs), MoBi’s ensure conservation resources are appropriately distributed.

**SPECIES DISTRIBUTION MODELLING**

However, conservation science is a crisis discipline and data collection is not able to keep up with rapid decision-making requirements. Old, limited, and incomplete data are frequently used as a ‘best guess’ in times of need. The challenge is even greater when data is required over particularly wide-ranging or long-running scales due to the limited resources available to biodiversity research. In the absence of consistently reliable and widespread observational data, scientists rely on methods to approximate distributions from smaller amounts of data. The most widely used are the aptly named species distribution models (SDMs).

SDMs are mathematical models that relate the observed distribution of a species to a set of environmental predictor variables (e.g., climate, habitat type, distance to water). Despite their widespread use in species and habitat conservation, these models contain a degree of uncertainty associated with the predictions.

**SOURCES OF INACCURACY**

The accuracy of any analysis or prediction depends on the quality of the original data. In a paper published in 2013, Joaquín Hortal (National Museum of Natural Sciences, MNCN-CSIC, Spain) and Richard Ladle (CIBIO-InBIO, University of Porto, Portugal) identified three major factors that compromise the quality of species distribution knowledge at any given spatial scale: survey completeness, the decay of information with time, and the decay of information with space. These decays account for the progressive loss of value of data to predict biodiversity trends at increasing distance from the observations.

Survey completeness can be impacted by the difficulty of detecting species due to their characteristics – such as cryptic colouration, elusive behaviour, or high mobility – and the characteristics of their habitat. As a result, even biological surveys employing a range of sampling methods can be incomplete. Hortal, this time working together with Juliana Stropp (MNCN-CSIC) and collaborators, found that sampling sites for butterflies in the Brazilian Atlantic Forest were biased towards large and connected forest fragments (Sobral-Souza et al, 2020).

**CONSERVATION BRAZILIAN ATLAS**

An SDM constructed with incomplete survey data like this, the inference of the functional relationship between butterflies and deforestation would be limited to these well-sampled areas. Increasing sampling effort in small and disconnected forest fragments would lead to more accurate evaluations of landscape-scale effects in the future that can be used by conservation decision-makers. A similar study on Brazilian mosses (Ronquillo et al, 2020) found that there were also spatial gaps and biases in the dataset and dates were lacking from the metadata of many records, further limiting analysis in temporal biodiversity shifts associated with global change stresses.

**BIOGEOGRAPHICAL IGNORANCE**

Biogeographical Ignorance (MoBi) that do just this – they account for survey completeness and quality of data in each area, as well as the decay in information (Tessarolo et al, 2021). Highlighting these uncertainties helps identify future conservation priorities. They can determine areas with more reliable predictions and support decision-makers using SDMs, for example in the allocation of future conservation priorities. Regularly updating the combined distribution and ignorance maps would provide users with the information required to deal with biogeographical uncertainty through the design and implementation of new surveys or by incorporating partially explicit estimates of uncertainty into future SDMs. A combined approach would allow conservation decision-makers to prioritise the existing conservation needs of species and allocate resources effectively, for example by focusing resources to survey and conserve the biodiversity of areas that are vulnerable to deforestation or climate change but remain poorly sampled.

**PROOF OF CONCEPT**

To prove the concept, the research group combined analytical and visualisation tools to represent four main sources of biological ignorance – data completeness, taxonomic quality of data, temporal decay in information, and spatial and environmental distance to the surveyed site – in a Scarabaeidae dung beetle dataset from the Iberian Peninsula. A biogeographical ignorance value was calculated for each record, generating a MoBi for each species. This ignorance value represents the degree of reliability of the biogeographical information for a species in a location.

Figure 1. Spatial coverage and bias of sites with well-sampled fruit-feeding butterfly inventories, in relation to the overall climate conditions in the Brazilian Atlantic Forest.
So the higher the value, the less reliable the information in the area is.

A biogeographical ignorance value for the focal species is attributed to each cell of the studied area depending on whether the cell contains original data on the focal species, a species from the studied taxonomic group, or neither of these. If the cell contains a record for the focal species, the biogeographical ignorance value is taken from the most reliable record. If the cell contains a record for the focal species, the biogeographical ignorance value is taken from the most reliable record. If the cell contains data for the studied species group, environmental and spatial proximity to cells with data is the only way to calculate ignorance. The researchers combined SMD maps with MoBIs to visually represent the reliability of distribution knowledge.

The study described four extreme (high or low) combinations of suitability and ignorance: 1) areas with high prediction suitability for the focal species to occur and high ignorance; 2) areas of high suitability and low ignorance; 3) areas with low predicted suitability and low ignorance; and 4) areas of low suitability and high ignorance. In terms of conservation priority based on these outputs, areas of high ignorance would benefit from further surveys, and areas of low ignorance could be targeted with conservation management to increase suitability or preserve habitat, if they are predicted to be home of the focal species.

Ignorance maps that are combined with SMD predictions generate spatially explicit maps of uncertainty and reveal insights into the reliability of such commonly used maps. Neglecting this uncertainty can affect SDM interpretation and thus potentially incorrect theoretical or practical applications, including ill-advised conservation actions. By routinely using maps of ignorance or similar techniques, conservation action can become more effective. In a resource-deficient discipline such as conservation science, it pays to understand the data and its inadequacies before making conservation decisions that could waste valuable resources or lose invaluable habitats and species.

References


Personal Response

What is next for your research group?

First, we have to adapt the parameters used to generate MoBIs to different case studies, extending their application to the particularities of as many types of species, communities and habitats as possible. In parallel, we will develop easy-to-use software tools to calculate MoBIs from biodiversity databases, and incorporate them into SMD software suites tailored to both model data and display its uncertainty. This will help integrate MoBIs as spatially explicit error terms into particular SDM applications, so they account for data-driven uncertainty. These novel tools will be used to assess the biogeographical ignorance in the tropical biomes of Africa and South America through TROPIBIO and TAXON-TIME projects, led by Richard Ladle and Juliana Stropp, respectively.