FIT3161/2 Computer Science Project and FIT3163/4 Data Science Project

Semester 1, 2019

Verification of species observation points for the Victorian Biodiversity Atlas using predictive modelling

project proposed by

Department of Environment, Land, Water and Planning DELWP

Version 1.0, 22 March 2019

Background

Species observation points (flora and fauna) are one of the most important factors when developing species distribution models for Victoria. These models are used for decision-making in landscape management (e.g. reducing invasive species https://www.environment.vic.gov.au/biodiversity/natureprint), fire management (e.g. guiding resources at preserving habitat of threatened species) and in issuing permits (e.g. allowing or forbidding certain activities to ensure the last harm to threatened species or habitats https://nvim.delwp.vic.gov.au/).

Species observations are entered by volunteers though an online application (https://www.environment.vic.gov.au/biodiversity/victorian-biodiversity-atlas/vba-go) that assures data completeness. New tools and citizen science contributions mean the number of observations contributed are rapidly growing. This is placing a burden on the expert verification processes as data validity currently needs to be confirmed by species experts before the data is stored in the observations repository known as Victorian Biodiversity Atlas (VBA) (https://www.environment.vic.gov.au/biodiversity/victorian-biodiversity-atlas/about-the-vba).

Goals

The main goal of this project is to develop a predictive model that sorts new observations into high and low reliability categories with the intention of keeping the Victorian Biodiversity Atlas as up to date as possible. The categorisation will guide how DELWP best uses their experts' time. The second goal is to identify predictors (variables) that have the most impact on successful categorisation.

Datasets

The DELWP Biodiversity Division is providing verified datasets of species observations with location, high and low reliability labels and collection method attributes for several selected species. There are several hundred observations per species (Figure 1).

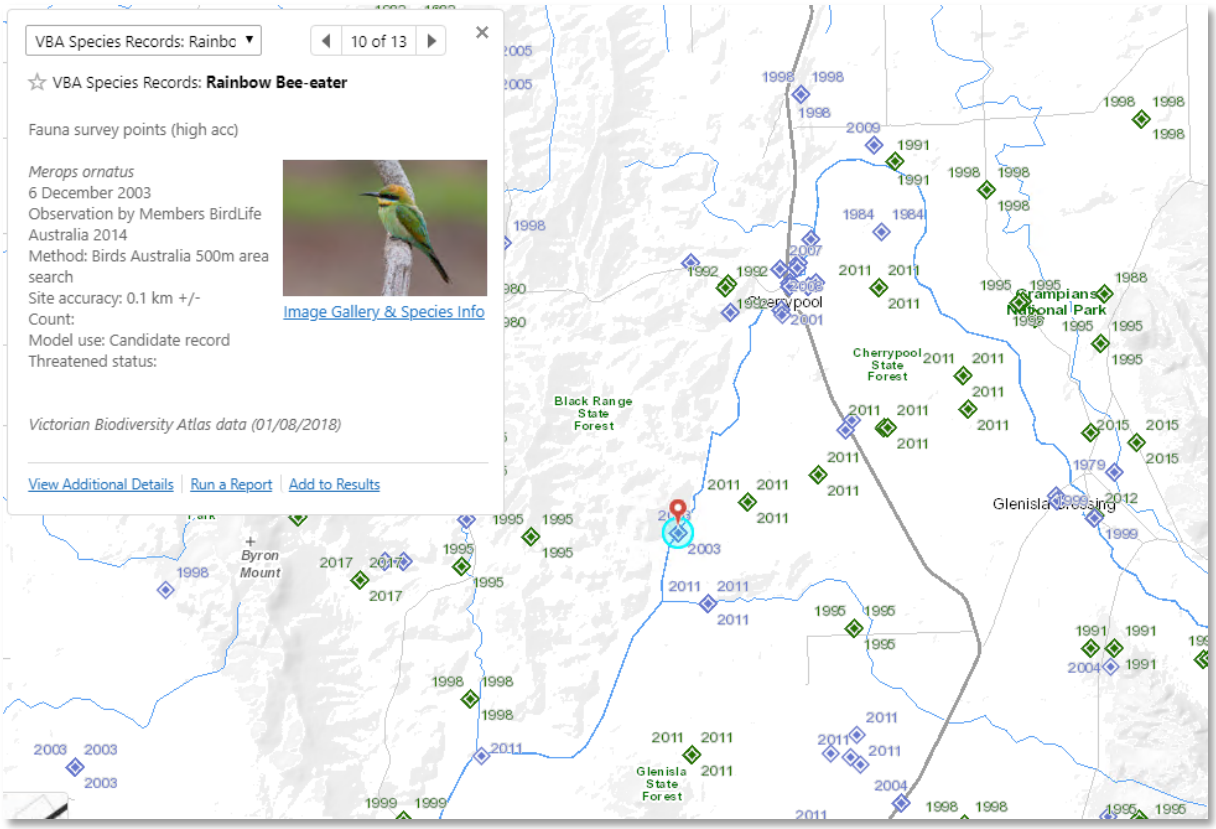

Figure 1. Observation example.

In addition, unverified observation datasets, and geospatial environmental datasets are provided and can be completed with online datasets. These dataset can be used as predictors for the classification model. Examples include vegetation type, ecoregions, elevation, radiometrics, wetness, rainfall, temperature, saline wetlands, Landsat satellite imagery, logging and fire history, etc.

After studying the habitat requirements of the species allocated to them, students should select a number of additional potentially relevant datasets from public environmental data repositories.

Method

Data preparation: Students need to link relevant environmental data (listed above, plus the self-selected datasets) to the verified observation locations, thus creating additional attributes for each observation point on top of the collection method attributes. Geospatial processing will be required for this step to extract attribute values for the observation points.

Modelling: Students will use different predictive modelling algorithms to categorise a training and a test dataset. The attributes extracted from the environmental datasets will function as predictors that replace the expert's specialist knowledge. Correct oversampling of training/test datasets, parameter exclusion due to high correlation, and possibly outlier exclusion must be considered.

Outcome

Report – Discussion of Method

The DELWP Biodiversity Division is interested in understanding and comparing the resulting models. Students should submit an evaluation of the modelling algorithms' suitability, including information on accuracy, sensitivity and specificity. Relevance of environmental attributes and their relationships should be discussed. A map including the already verified plus the newly categorized high reliability data could be created for the species.

System functionality

The DELWP Biodiversity Division is keen on running the resulting models. Students should submit the source code and compiled code as well as a documentation of the code, explaining how to rerun the model including tweaking model parameters. Ideally, student groups would bring their models together in a common framework that selects the individual model based on the species name of the entered observation.

Users at DELWP will use the system via a command line tool for processing observation points in batch mode. A REST web service with a graphical user interface is also of interest. If they wish, students can automate the geographic analysis to enrich observations points with additional attributes. This would require setting up and adapting a geographic information system with custom scripts. More information about these and other related aspects regarding system functionality will be provided at a later time.

Programming language and code

DELWP prefers platform independent frameworks and programming languages, such as Python and JavaScript. Java code is also acceptable, but C/C++ and other similar languages should be avoided if possible.

References

For a useful list of literature, please see https://www.ari.vic.gov.au/research/modelling/habitat-distribution-models-hdms