Thesis Outline

Lukas Prader

1   Introduction

(citations missing)

The topic of invasive species has become more and more important, even more so with recent changes in climate and habitats due to human influence. In order to deal with invasive species and the impact they can have on existing ecosystems, species distribution models (SDMs) have been used to predict the potential habitat and through that the threat of emerging invasive species. Especially time-partitioned models can provide more insight into the process of invasion.

Harmonia axyridis, also kown as harlequin or Asian ladybeetle, is an invasive species already established in non-native habitats all around the world. The invasion process has already been studied extensively and there have been SDMs modelling the invaded range as well. This thesis aims to analyse the spread of *Harmonia axyridis* in detail, as well as creating new insight into the progression of an ensemble SDM iterating over each year.

2 Methods

(citations missing)

2.1 Datasets

In order to also have some kind of influence of the human nature of first introduction, land cover data was used from the Copernicus Land cover Classification dataset with yearly resolution starting from 1992 up to 2020. All traditional 19 bioclim variables were obtained from the CHELSA V2.1 climatologies datasets too, using the 1981-2010 time frame for all years from 1992 to 2010 as well as the MPI-ESM 1.2 ssp370 scenario 2011-2040 for all years from 2011 to 2022. For occurence data, all global occurences of *Harmonia axyridis* were downloaded from the Gbif database.

2.2 Data preparation

All bioclim and land cover layers were resampled to a matching resolution of 30 arc seconds and cropped to two spatial extents, europe and the presumed native range referencing (Orlova-Bienkowskaja, Ukrainsky & Brown, 2015). The presence-only points from Gbif were subset to the afore mentioned spatial extents and then checked for missing values for latitude, longitude, year or coordinate uncertainty. No occurences after 2022 were used, also no points with a coordinate uncertainty larger than 1 km. Afterwards, using the library CoordinateCleaner, all remaining datapoints were again checked for common errors or biases in the respective subset. To prepare the data for modelling, pseudoabsences were generated for each year. For this, the subset of a specific year was taken and used to generate pseudo-absences limited to a radius around the occurence points (Phillips et al. 2009). (correction for overlap with following years?)

2.3 Model building

For each year, the following Models were computed: General Linear (GLM), General Additive (GAM), Multivariate Adaptive Regression Splines (MARS), Boosted Regression Trees (BRT) and Maximum Entropy (MAXENT). (Model choices subject to change) Each model used either only datapoints from europe (eu) or from europe and the native range in asia (eu+as). A model for a specific year allways included all points from past years as well.
To select the variables used, bioclim variables were chosen on . . . For the land cover data, a PCA was computed and added to the chosen bioclim variables.

2.4 Analysis

All SDM models of each year were tested for their accuracy on predicting the occurences of the following year and the final year of 2022. For each year, the occupied niche was also computed, though only using datapoints from the year in question (Broennimann et al. 2011). These niches were again compared to the following year and 2022.