Thesis

Lukas Prader

5th November 2023

Contents

1 Introduction

Species Distribution Models (SDMs) attempt to connect species distribution data, i.e. occurrence at a location, to the environmental and spatial properties of said location[1]. They have been shown to generate substantial insight into the ecological requirements of species and, as niche models, can be used to predict the potential habitat of a species[2]. There has been considerable debate on the capabilities and limitations of SDMs, especially when used for prediction. In general, SDMs are made with the (ideal) assumption that the species is in environmental equilibrium[1], implying that its ecological niche is not currently changing. If these models are now used to predict new, unsampled areas, there actually is no measure to assess their accuracy, since no data is presently available for that area[2]. There is also no guarantee that the biotic interactions sampled in the study area will reflect the final interactions in the new area[1]. All of these issues apply especially to the prediction of invasive species, since there might be limited data in the invaded range, the species is often not currently at equilibrium and interactions with native species are completely new[3]. Despite all these challenges, SDMs have been used numerous times to provide insight into the invasive potential and the invasion dynamics of alien species[4]. One way of gaining more insight into the invasion process is to create models with data from different time periods during the invasion[5]. For example, data from a time period early in the invasion process can be used to build models which then are evaluated against data from a later time period[6].

A multitude of Modelling approaches have been developed to model ecological processes. In SDMs, methods range from regression methods to machine learning and each feature various strengths and weaknesses, possibly leading to vastly different results for the same dataset. Due to those differences, a possible approach is to create an ensemble of multiple models[7]. The way of combining model predictions can vary, but the goal is to improve total performance by combining the results of all computed models.

Invasive species are of special interest in ecological research due to their impact on native ecosystems. Main goals in research are to find out which species have potential to become invasive, what habitat will be susceptible to invasion by those species, how fast the species will invade the new area and what impact its invasion will have on the native ecosystem[8]. To this end, many theories have been created to describe invasion processes. The invasion of a species can generally be described with four stages[9]:

1. Transport: Leaving the native range, arriving at a new location

2. Introduction: Existing in specific locations (captivity / cultivation)

3. Establishment: Existing outside of areas of introduction in the wild

4. Spread: Sustaining establishment and dispersing to new environments

The impacts of invasive alien species can be numerous, ranging from food web changes to reductions in habitat and species richness, hydrology and nutrient cycle changes, enhanced invasion of other species and mass extinctions[10]. For example intraguild predation, the predation of species using similar resources, can create completely new stable states of an ecological system[11].

Harmonia axyridis, also known as the Harlequin ladybird or multicoloured Asian lady beetle, is of the family of the Coccinellidae and has its native origin in Asia[12]. At first widely introduced as a control species against pest aphids, *H. axyridis* has turned out to be a highly invasive species reaching an almost global distribution[13]. In America, the species was introduced as early as 1916 (California) and in 1988, first populations outside intended release were found[14]. Usage of *H. axyridis* for biological control in Europe dates back as far as 1990 (France)[15]. First invasive occurrences were confirmed in multiple countries during the early 2000s, including Germany (2000), Belgium (2001), the Netherlands (2002) and the United Kingdom (2003)[12]. The first confirmation in Austria, where it was never used for biological control, was 2006[16]. It has been shown that all established invasive populations outside of North America have their origin in the first established population in eastern North America, with the European populations being significantly influenced by the used biocontrol strain[17].

The impact of *H. axyridis* on invaded areas is diverse. In some contexts, the ladybird has been shown to have a negative impact on the diversity and abundance of native ladybird species[12]. Many studies show intra guild predation and direct interspecific competition in favour of *H. axyridis*[18]. It has also been shown that the species feeds on a variety of damaged fruit crops, for example grapes, apples, stone fruit and berry crops[19]. The aggregating behaviour of *H. axyridis*, mostly as a strategy for overwintering, is also a cause of disturbance, since private homes and facilities are invaded by large amounts of beetles at a given time[20].

One goal of this thesis is to look into the limitations of models built early in the invasion process of a species. By iterating over the years of the invasion, model performance can be evaluated with consideration to the current state of invasion. Computing the occupied niche separately for each year also

provides more insight into the invasion process and might help to understand modelling limitations for certain years.

2 Methods

This section elaborates on the Methods used to conduct this research.

2.1 Datasets

For occurrence data, all global occurrences of *H. axyridis* were downloaded from the GBIF database[21]. All traditional 19 bioclim variables were obtained from the CHELSA V2.1 climatologies dataset[22,23], using the 1981-2010 time frame for all years from 2002 to 2010 as well as the MPI-ESM 1.2 ssp370 scenario 2011-2040 for all years from 2011 to 2022. 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[24] with yearly resolution for 2002 up to 2020.

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)[25].

The presence-only points from GBIF were checked for missing values for latitude, longitude, year or coordinate uncertainty and then subset to the afore mentioned spatial extents. No occurrences after 2022 were used, also no points with a coordinate uncertainty larger than 1 km. In Europe, the initial cut off year for presences was 1991, since this is the year of invasion according to the EASIN website. Afterwards, using the library `CoordinateCleaner`, all remaining data points were again checked for common errors or biases in the respective subset (tests used: "capitals", "centroids", "duplicates", "equal", "institutions", "outliers", "seas", "zeros"). In addition, all occurrences were checked for their land cover class values in their respective year, removing points in the water or with no data. In the end, remaining data points prior to 2002 were deemed insignificant and removed from the dataset. To prepare the data for modelling, pseudoabsences were generated for each year, randomly sampling the area and resampling points in the water or with no data.

2.3 Model building

For each year, the following Models were computed: General Linear (GLM)[26], General Additive (GAM)[26], Boosted Regression Trees (BRT)[27] and Maxi-

mum Entropy (MAXENT)[28]. A model for a specific year always included all points from past years as well. The iterative models that were built only use data points from Europe, though there was one model created only with native occurrences and predicted for each year in Europe. For all used occurrence points after 2020, the land cover data of 2020 was used as a substitute.

Variance inflation factors were used to select the variables used for model building. For this, a GAM was computed only using Europe data from 2002, using all bioclim and land cover variables. For land cover variables, a PCA was computed on the relative area of all land cover classes in an 18 km radius around 5000 random data points in Europe. The 18 km radius was chosen, since it is the average flight distance determined for *H. axyridis*[29]. PCA axes were included in the model until a cut-off of 80% of explained variance was reached. Variance inflation factors were computed for this GAM and the variable with the highest VIF was dropped until none of the remaining variables had a VIF greater than 10.

2.4   Analysis

All SDM models of each year were evaluated for their accuracy on predicting the occurrences of the following year and the final year of 2022 using the True Skill Statistic (TSS). The TSS values were used to create a TSS-weighted ensemble of all model predictions, which was again evaluated for its accuracy. For each year, the occupied niche was computed by running a PCA analysis on the bioclim variables. The niche can then be visualized by plotting a dynamic occurrence density grid for the first two PCA axes[30]. The overlap between each year for Europe and the respective following year was computed, as well as a niche similarity and niche equivalency test. Niche overlap for each year was also visualized in comparison to the total native niche. All mentioned niche analyses were conducted using the library `ecospat`[31].

References

1. Elith, J. & Leathwick, J. R. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics* **40,** 677–697. doi:`10.1146/annurev.ecolsys.110308.120159` (2009).

2. Araújo, M. B. & Guisan, A. Five (or so) challenges for species distribution modelling. *Journal of biogeography* **33,** 1677–1688. doi:`10.1111/j.1365-2699.2006.01584.x` (2006).

3. Mainali, K. P. *et al.* Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global change biology* **21,** 4464–4480. doi:`10.1111/gcb.13038` (2015).

4. Zimmermann, N. E., Edwards Jr, T. C., Graham, C. H., Pearman, P. B. & Svenning, J.-C. New trends in species distribution modelling. *Ecography* **33,** 985–989. doi:`10.1111/j.1600-0587.2010.06953.x` (2010).

5. Briscoe Runquist, R. D., Lake, T., Tiffin, P. & Moeller, D. A. Species distribution models throughout the invasion history of Palmer amaranth predict regions at risk of future invasion and reveal challenges with modeling rapidly shifting geographic ranges. *Scientific Reports* **9,** 2426. doi:`10.1038/s41598-018-38054-9` (2019).

6. Barbet-Massin, M., Rome, Q., Villemant, C. & Courchamp, F. Can species distribution models really predict the expansion of invasive species? *PloS one* **13,** e0193085. doi:`10.1371/journal.pone.0193085` (2018).

7. Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends in ecology & evolution* **22,** 42–47. doi:`10.1016/j.tree.2006.09.010` (2007).

8. Shigesada, N. & Kawasaki, K. *Biological invasions: theory and practice* `https://books.google.at/books/about/Biological_Invasions_Theory_and_Practice.html?id=Ri-hle_zdpsC&redir_esc=y` (Oxford University Press, UK, 1997).

9. Blackburn, T. M. *et al.* A proposed unified framework for biological invasions. *Trends in ecology & evolution* **26,** 333–339. doi:`10.1016/j.tree.2011.03.023` (2011).

10. Simberloff, D. *et al.* Impacts of biological invasions: what's what and the way forward. *Trends in ecology & evolution* **28,** 58–66. doi:`10.1016/j.tree.2012.07.013` (2013).

11. Polis, G. A., Myers, C. A. & Holt, R. D. The ecology and evolution of intraguild predation: potential competitors that eat each other. *Annual review of ecology and systematics* **20,** 297–330. doi:`10.1146/annurev.es.20.110189.001501` (1989).

12. Roy, H. E. *et al.* The harlequin ladybird, Harmonia axyridis: global perspectives on invasion history and ecology. *Biological invasions* **18,** 997–1044. doi:`10.1007/s10530-016-1077-6` (2016).

13. Brown, P. *et al.* Harmonia axyridis in Europe: spread and distribution of a non-native coccinellid. *BioControl* **53,** 5–21. doi:`10.1007/s10526-007-9132-y` (2008).

14. Chapin, J. B., Brou, V. *et al.* Harmonia axyridis (Pallas), the third species of the genus to be found in the United States (Coleoptera: Coccinellidae). *Proc. Entomol. Soc. Wash* **93,** 630–635. `https://www.biodiversitylibrary.org/partpdf/55539` (1991).

15. Coutanceau, J. Harmonia axyridis (Pallas, 1773): Une coccinelle asiatique introduite, acclimatée et en extension en France. *Bulletin de la Société Entomologique de France* **111,** 395–401. doi:`10.3406/bsef.2006.16343` (Jan. 2006).

16. Rabitsch, W. & Schuh, R. First record of the multicoloured Asian ladybird Harmonia axyridis (Pallas, 1773) in Austria. *Beiträge zur Entomofaunistik* **7,** 161–164. `https://www.zobodat.at/pdf/BEF_7_0161-0164.pdf` (2006).

17. Lombaert, E. *et al.* Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PloS one* **5,** e9743. doi:`10.1371/journal.pone.0009743` (2010).

18. Pell, J. K., Baverstock, J., Roy, H. E., Ware, R. L. & Majerus, M. E. Intraguild predation involving Harmonia axyridis: a review of current knowledge and future perspectives. *BioControl* **53,** 147–168. doi:`10.1007/978-1-4020-6939-0_10` (2008).

19. Koch, R., Burkness, E., Burkness, S. W. & Hutchison, W. Phytophagous preferences of the multicolored Asian lady beetle (Coleoptera: Coccinellidae) for autumn-ripening fruit. *Journal of Economic Entomology* **97,** 539–544. doi:`10.1093/jee/97.2.539` (2004).

20. Nalepa, C., Kennedy, G. & Brownie, C. Role of visual contrast in the alighting behavior of Harmonia axyridis (Coleoptera: Coccinellidae) at overwintering sites. *Environmental Entomology* **34,** 425–431. doi:`10.1603/0046-225X-34.2.425` (2005).

21. Gbif.org. *GBIF Occurrence Download* 15th July 2023. doi:`10.15468/dl.y4y3s9`.

22. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Scientific data* **4,** 1–20. doi:`10.1038/sdata.2017.122` (2017).

23. Karger, D. N. *et al. Climatologies at high resolution for the earth's land surface areas* (Last accessed 3 August 2023). 2021. doi:`10.16904/envidat.228`.

24. Copernicus Climate Change Service. *Land cover classification gridded maps from 1992 to present derived from satellite observation* (Last accessed 3 August 2023). 2019. doi:`10.24381/cds.006f2c9a`.

25. Orlova-Bienkowskaja, M. J., Ukrainsky, A. S. & Brown, P. M. Harmonia axyridis (Coleoptera: Coccinellidae) in Asia: a re-examination of the native range and invasion to southeastern Kazakhstan and Kyrgyzstan. *Biological Invasions* **17,** 1941–1948. doi:`10.1007/s10530-015-0848-9` (2015).

26. Guisan, A., Edwards Jr, T. C. & Hastie, T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling* **157,** 89–100. doi:`10.1016/S0304-3800(02)00204-1` (2002).

27. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *Journal of animal ecology* **77,** 802–813. doi:`10.1111/j.1365-2656.2008.01390.x` (2008).

28. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. Opening the black box: An open-source release of Maxent. *Ecography* **40,** 887–893. doi:`10.1111/ecog.03049` (2017).

29. Jeffries, D. L. *et al.* Characteristics and drivers of high-altitude ladybird flight: insights from vertical-looking entomological radar. *PloS one* **8,** e82278. doi:`10.1371/journal.pone.0082278` (2013).

30. Broennimann, O. *et al.* Measuring ecological niche overlap from occurrence and spatial environmental data. *Global ecology and biogeography* **21,** 481–497. doi:`10.1111/j.1466-8238.2011.00698.x` (2012).

31. Di Cola, V. *et al.* ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40,** 774–787. doi:`10.1111/ecog.02671` (2017).

Appendix