ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 11/07/2025

DOI: https://doi.org/10.3897/arphapreprints.e164783

Optimizing invasive species distribution models: a test with the giant hogweed in Switzerland

Bart Steen, Antoine Guisan, Luigi Maiorano, Laura Celesti-Grapow

Optimizing invasive species distribution models: a test with the giant hogweed in Switzerland

Bart Steen*[1,2], Laura Celesti-Grapow[3], Luigi Maiorano**[2], Antoine Guisan**[1,4]

[1] Institute of Earth Surface Dynamics, University of Lausanne, CH-1015 Lausanne Switzerland
[2] Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Roma, Italy
[3] Department of Environmental Biology, Sapienza University of Rome, Piazzale A. Moro 5, 00197 Roma, Italy
[4] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne Switzerland

* Corresponding author
** Shared last authorship

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 11/07/2025.* DOI: https://doi.org/10.3897/arphapreprints.e164783

Abstract

Invasive alien species are one of the primary drivers of biodiversity loss and cause great economic damage worldwide. They can be managed, but anticipating their invasions is paramount. This can be accomplished by modelling their current and future distributions. Species Distribution Models are a staple tool for this purpose and recent advances to use them to model invasive alien species have been made and are applied here. In this paper, two common species distribution modelling methods, MaxEnt and Random Forests, are employed to simulate the present and future potential distributions dangerous invasive species, *Heracleum mantegazzianum* Sommier and Levier, using opportunistically collected occurrence datasets, a nested modelling approach and background points sampled stratified randomly in environmental space. Possible bias in the occurrence datasets is treated in several ways and the differences in output distributions are discussed, as well as the uses of the outputs for guiding eradication efforts.

30 Keywords

31 Invasive alien species ; *Heracleum mantegazzianum* Sommier and Levier ; SDMs ; Ecological niche
32 modelling ; background points ; environmental space ; Sampling bias ; future distributions ; hierarchical
33 approach ; spatially-nested approach

Introduction

Invasive alien species (IAS) are species that are causing economic or natural damage and are exotic in parts of their ranges, causing an estimated $391 billion economic damage per year (IPBES 2023) and posing a great threat to biodiversity (UNEP 2023). Their spread has been greatly facilitated by world trade and human movements. Moreover, ongoing climate change increasingly allows them to access areas that have remained mostly invasion free until recently, like the polar regions and high mountains (Pauchard et al. 2009, Petitpierre et al. 2016, Haider et al. 2022).

Species Distribution Models (SDMs) (Franklin 2010; Guisan et al. 2017) are becoming increasingly more advised as a tool in conservation planning and management (Guisan et al. 2013) but often remain underused in practice (Cassini 2011; Tulloch et al. 2016; Steen et al. 2024a). Recent advancements on creating standards for SDMs (Araujo et al. 2019) and producing accurate predictions of species distributions have been made (Kass et al. in press) and need to be applied in practice to create accurate distribution maps for practitioners, a step that is especially paramount for monitoring IAS (Lake et al. 2020).

For instance, it was determined that sampling background points stratified randomly in environmental space yields the best performance metrics (Steen et al. 2024b). In addition, the use of a spatially-nested approach was shown to be crucial for modelling IAS (Gallien et al. 2012; Adde et al. 2023). The reason is, an optimized SDM for the purposes of management would ideally make use of high-resolution data and cover the entire species range, so that all known environmental conditions in which the species is known to exist are included in the model. Unfortunately, such high-resolution data is only available at local and regional scales, making the task of creating accurate SDMs for the purposes of conservation efforts harder. A spatially-nested approach allows both capturing the whole realized climatic niche over the entire species range at coarse resolution and large extent and fitting the species' fine environmental requirements at high resolution in a smaller extent. This allows overcoming the risk of niche truncation (Chevalier et al. 2021; Adde et al. 2023; Guisan et al. 2025) and ensures that all known conditions for the species are accounted for in the case of niche shifts between the native and invaded ranges (Broennimann and Guisan 2008), which can affect IAS (Guisan et al. 2014; Fernandez and Hamilton 2015; Bates et al. 2020; Aravind et al. 2022). Moreover, species presence is likely predicted by different environmental variables at different scales (i.e.,

spatial extents and resolutions), e.g. climate being hypothesized to be more important at large resolution and extent while soil or land cover being hypothesized to be more important toward finer resolutions (Pearson and Dawson 2003; Dawson 2004; Ni et al. 2007; McGill 2010; Vicente et al. 2014). The nested approach can link the largest scale to the smallest scale, thus allowing for selection of different environmental variables at different extents (Guisan et al. 2025).

Another potential drawback of most SDMs is that they are fitted using citizen science datasets, which often suffer from spatial and temporal sampling bias (Bird et al. 2014; Callaghan et al. 2019; Anderson et al. 2020). However, the exact nature of this bias is hardly possible to determine (Di Cecco et al. 2021). Therefore, any efforts to correct the bias by, for instance, subsampling the data might also eliminate valuable ecological information (Steen et al. 2021), making it hard to determine the appropriate treatment of the occurrence data (if any).

Heracleum mantegazzianum is an extensively studied IAS (Pyšek et al. 2007; Shackleton et al. 2020). It causes severe health hazards due to its photoreactive poison (Bhowmik and Chandran 2015) and obstructs ecosystem services like recreational fishing (Caffrey 1999; Nasadiuk and Mamchur 2024). In addition, its ability to form dense monocultures and alter the nutrient environment makes it a severe threat to biodiversity (Caffrey 1999). It invades areas left open by clearcuts and abandoned agricultural ground, mesic grasslands, roadsides, forest edges and eutrophied and iron-polluted soils forming tall-herb stands (Pyšek et al. 2007; Thiele and Otte 2007). For these reasons, *H. mantegazzianum* is both on the List of Union Concern (European Commission 2024) and on the Swiss blacklist of IAS (Infoflora.ch).

The species is spread mainly by humans (Bhowmik and Chandran 2015). It attaches itself to human clothes and shoes, and therefore, spreads through any vehicle humans use. Its natural reproduction is through wind-borne spores (Müllerová 2024). It is a low to mid elevation mountain species originally from the Caucasus mountains and displays generalist behavior in its invaded range. In Germany, it favors abandoned agricultural areas, likely profiting from high nutrient content of soils and absence of land management (Thiele et al. 2007; Thiele and Otte 2008; Thiele et al. 2008). In Europe, it is already found at altitudes above 1500 m (Shackleton et al. 2020; Müllerová 2024). It will probably follow the trend of invaders to spread to still higher elevations (Pauchard et al. 2009; Petitpierre et al. 2016), where it can potentially disturb

the many hotspots of alpine biodiversity, which provide invaluable ecosystem services (Grêt-Regamey and Kytzia 2007; Ramel et al. 2020).

In this study, we aim at using novel clues on optimizing SDM modelling approaches (Steen et al. 2024a; Steen et al. 2024b) to build the most optimized SDM for the potential present and future distributions of the IAS *H. mantegazzianum*.

We gathered the most recent findings on building accurate SDMs (e.g. Araujo et al. 2019), making use of multiple treatment strategies, and test them on the giant hogweed. In particular, we implemented a spatially-nested approach (Chevalier et al. 2021; Guisan et al. 2025) focusing on 3 different spatial extents: i) Global (i.e. all biomes where *H. mantegazzianum* is present), ii) Europe and iii) Switzerland. We also employed different classes of SDM algorithms and sampled background points in an environmentally random-stratified way (Steen et al. 2024b). In addition, we ran extensive model tuning and overfitting analyses, as well as different methods to deal with sampling bias. Finally, since it is not only important to know the current distribution of invasive species, but also its future distribution, we also projected the models in future climates.

Materials and Methods

Overall analytical framework

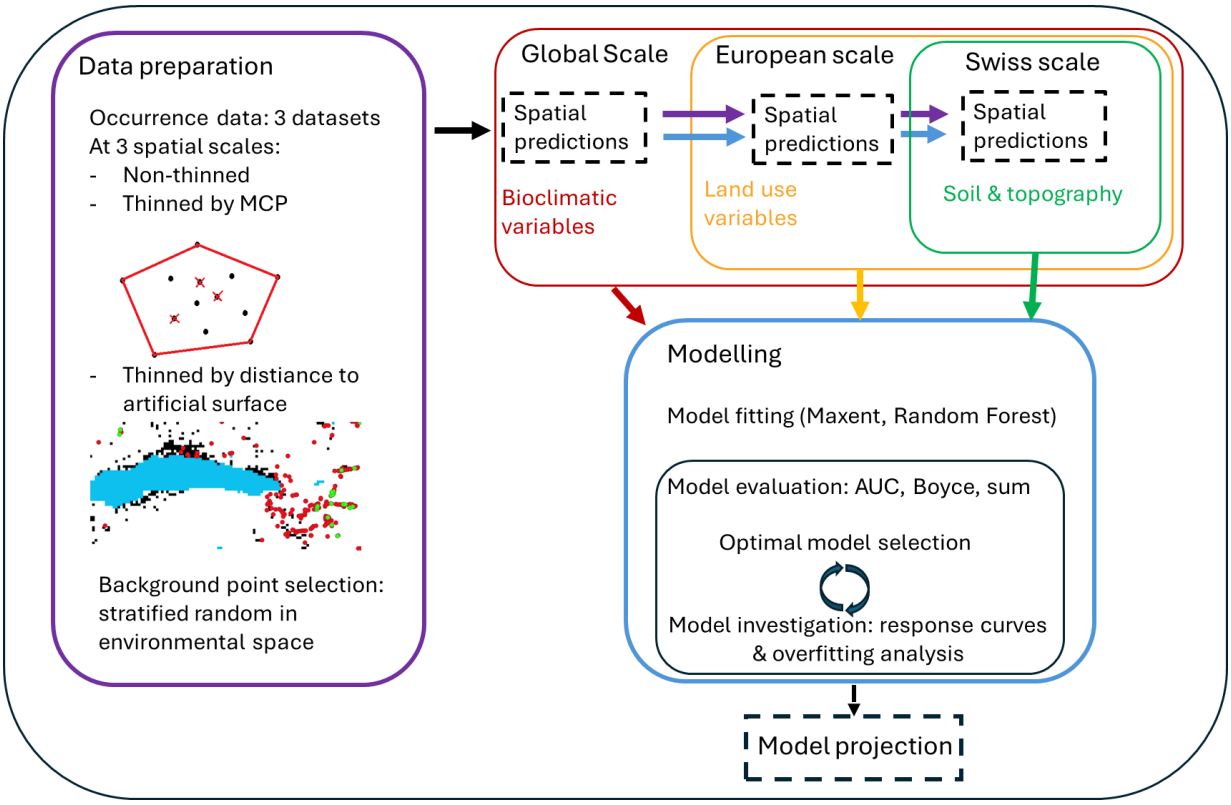

***Figure 1**: Overall analytical framework of the analyses performed.*

Environmental variables

Global bioclimatic variables were downloaded from CHELSA, version 2.1 (Karger et al. 2017), at a resolution of 0.0083 degrees (30 arc seconds); see Table 1. All variables are calculated as the average values from the time series of 1981–2010. Karger et al. (2017) provide also multiple projections of the CHELSA 2.1 variables according to several scenarios. In this study, the SSP 370 pathway was chosen for climatic projections.

At European scale, land use variables were used because in Central Europe, *H. mantegazzianum* grows mainly on anthropogenically disturbed areas, nutrient rich soil and abandoned agricultural ground (Thiele et al. 2008; Pergl et al. 2011). For present-day land cover variables, we used the 2015 datasets provided by the European Space Agency (Global ESA CCI land cover classification map), which were likewise at 30 arc seconds; see Table 1. Binary variables included were swampy and often flooded vegetation, mostly cropland and grass, shrub and

scrubland. In addition, the gridDistance() function from the R raster package was used to calculate the distance to forest, water, snow and bare/sparsely vegetated areas. Finally, the world population density (cropped to Europe) (Earth Science Data Systems 2024) was inserted as a variable.

At Swiss scale, environmental layers on soil and topography at 25 m resolution were used (See Table 1). They were taken from the SWECO25 dataset (Külling et al. 2024) and from the Swiss Federal Office of Agriculture (FOAG 2012).

We employed the Ecospat R package (Di Cola et al. 2017) to assess collinearity, at each spatial extent, among all these environmental variables. A threshold of 0.70 for correlation between variables was established, in accordance with Dormann et al. (2013). When two variables had a correlation >0.7, we kept only one of them. The definitive roster of variables chosen for model calibration across all three tiers is outlined in Table 1. For an ecological justification of the inclusion of each variable, see Appendix 2. The biomes where *H. mantegazzianum* does not occur, as well as all water bodies, were removed from all environmental variables.

153 ***Table 1**: Names, resolutions, units and descriptions of variables used at each spatial scale*
154 *(i.e., spatial extent) to run the Maxent and Random Forest models.*

Variable name	Unit	Description
World scale (30 arc seconds)		
bio2_V.2.1	°C	Annual temperature range
bio12_V.2.1	kg m^-2 year-1	Annual precipitation amount
bio15_V.2.1	kg m^-2	Precipitation seasonality
bio19_V.2.1	kg m^-2 month^-1	Mean monthly precipitation amount of the coldest quarter
Surface wind mean	m/s	Average surface wind velocity
Surface wind max	m/s	Maximum surface wind velocity
Growth degree days	number of days	Annual growth days - days with temperatures above 0 degrees
Growth season precipitation	kg m^-2 gsl^-1	precipitation sum accumulated on all days during the growing season. Gsl = growth season length in days
Growth season temperature	°C	Mean temperature of the growing season
Net primary productivity	g C m^−2 yr^−1	Net primary production
Europe scale (30 arc seconds)		
Bioclimatic envelope	N/A	Sum of thresholded world-scale models cropped to Europe scale
Distance to surface water	m	Distance to water. Calculated for every point on land inside the study area
Distance to sparse vegetation	m	Distance from bare or sparse vegetated areas. *H. mantegazzianum* was assumed to not occur in sparse vegetations.
Distance to snow and ice	m	-
Swampy or often flooded vegetation	N/A	-
Mostly cropland	N/A	Vegetation largely consisting of cropland
Grass, scrubs and shrubs	N/A	-
Distance to forest	m	Distance from nearest forest pixel (value = 0 if pixel is forest)
Worldpop_2000	people/km^2	Population density
Swiss scale (25m)		
Europe model		Sum of thresholded Europe-scale models cropped to Swiss scale
Digital elevation model (DEM)	m	Average elevation per 2x2 m square
Aspect	Degrees	Orientation of slope, averaged over 2x2 m squares, varying from 0 to 360

Slope	Degrees	Slope averaged over 2x2 m squares
Hillshade	Index (shaded to bright)	Measure of irradiation per 2x2m square.
Nutrient holding capacity	Index (1-6)	Ordinal measure of soil capacity to hold nutrients
Kultur	Index (1-5)	Ordinal measure of land productivity

Occurrence data

Data were collected from GBIF (GBIF 2021) and from an unbiased occurrence dataset collected using an opportunistic scientific sampling scheme (Celesti-Grapow et al. submitted). Only occurrence records from the year 2000 or later were selected, in order to match the CHELSA version 2.1 time series. We assumed that the time difference between 1981–2010 and 2000–2021 was representative of the delayed response of the species to shifting climate (Steen et al. 2024b). From the full database we eliminated improbable coordinates, occurrences characterized by coordinate uncertainty exceeding 50 meters, observations reporting zero abundance, and the removal of entries denoting an occurrence status of *absent* across multiple languages. Additionally, a visual inspection, checking for occurrence records where they should not be, was conducted. These data cleaning operations were executed as per Steen et al. (2024b). The three datasets were all used to run the SDMs at all three spatial extents.

Bias correction and data thinning

Given the different sampling designs of the individual datasets, three different versions of the full combined dataset were used. Firstly, a fully non-thinned dataset (so the entire dataset, except for selecting only 1 occurrence record per environmental variable raster grid), a MCP thinned dataset and a dataset based on the distance to artificial surface and urban area. These treatments were chosen because of the differences in sampling design and the different sampling biases. Secondly, an MCP-based thinning method (Steen et al. 2024b) was applied, which counters spatial autocorrelation but also might eliminate ecologically relevant information (Steen et al. 2021). Therefore, the third dataset was thinned based on the assumption that sampling intensity in citizen science databases is higher by human infrastructure. On the other hand, higher abundances of *H. mantegazzianum* along roads and cities might be realistic since humans are a major vector in spreading the species. Due to this uncertainty, we also used the entire, non-thinned occurrence point dataset. In the spatially nested approach, only SDMs fitted

using the same occurrence dataset treatment at larger scale were included in the lower scale models using that treatment.

The occurrence data were thinned as per Steen, et al. (2024b) as follows:
We delineated a Minimum Convex Polygon (MCP; Mohr 1947) encompassing the data points. This serves to define the boundaries of the species' home range. Subsequently, within this MCP, we implemented a random point selection method. This procedure was iterated 50 times, and on each iteration, we computed the average nearest neighbor distance between the randomly selected points. The overall average of these individual averages was then calculated. To reduce sampling bias and align the distribution of species occurrence records with what would be expected in a random distribution, we employed the spThin R package (Aiello-Lammens et al. 2015; Chiocchio et al. 2021).

The data thinned based on distance to artificial surface and urban area were created by extracting the “artificial surface and urban area” from the present-day land use data (Global ESA CCI land cover classification map) and calculating the distance from for each pixel that was not artificial surface or urban area using the gridDistance() function from the R raster package. The values of this raster dataset were then used as a “weight” to sample occurrence records from the non-thinned dataset using the sample() function in R. The number of occurrence points sampled was the same as the number of points remaining after the MCP thinning.

Both thinning steps were repeated at all 3 spatial extents, generating 6 thinned datasets. Since the non-thinned occurrence datasets were also used, a total of nine occurrence datasets have been used to generate the models (three at each spatial extent). The ones built for the Swiss scale are given in Figure 2.

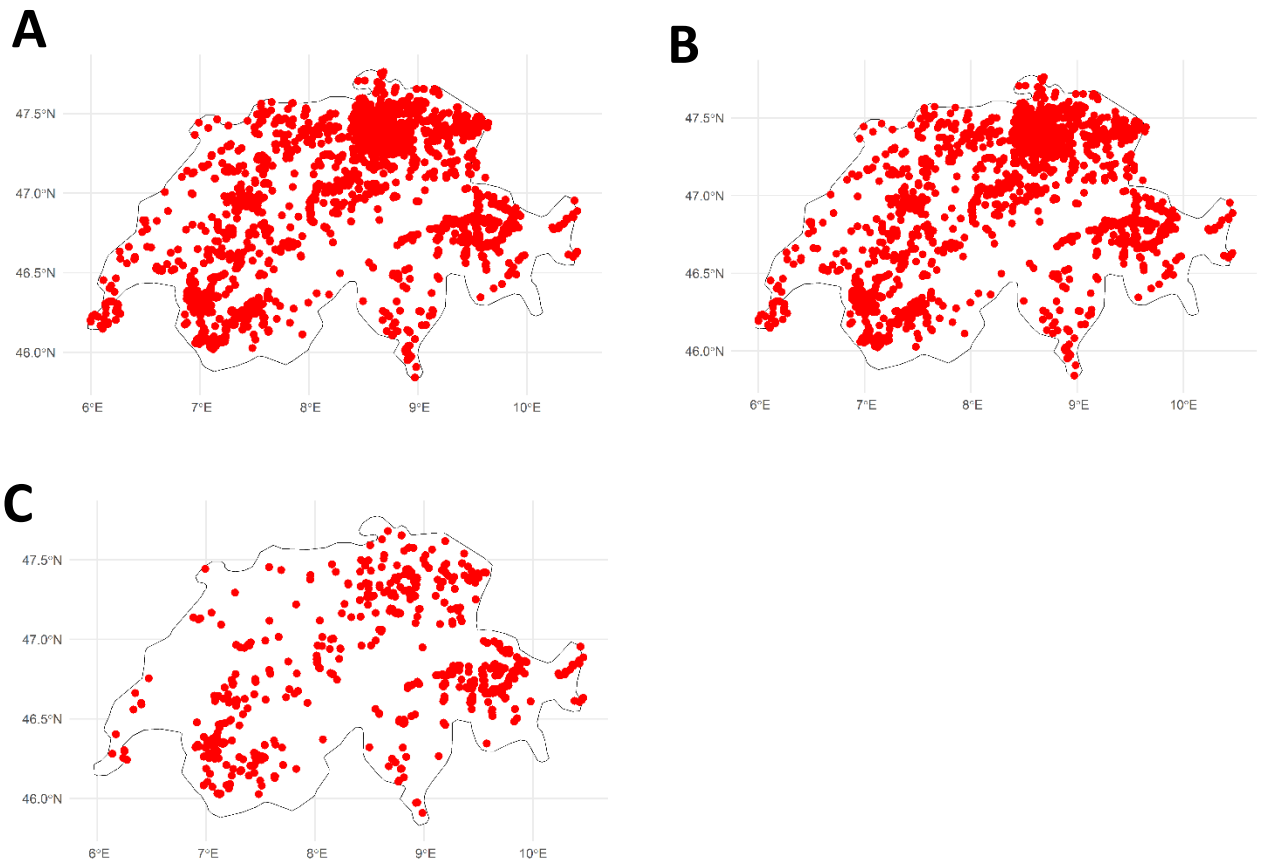

***Figure 2**: The three occurrence datasets used to fit the SDMs at Swiss scale: no thinning (A), thinned with MCP (B) and thinned by proximity to artificial surface and urban area (C).*

Species Distribution Models

Spatially-nested approach

The spatially-nested approach was implemented by running Species Distribution Models (SDMs) using the Maxent algorithm (Phillips et al. 2017) and Random Forests (standard implementation, using the randomForest() function from the randomForest R library (Breiman et al. 2024)). Models were run at three spatial extents: (i) global (from which all the biomes where *H. mantegazzianum* can occur were extracted), (ii) Europe (also with only *H. mantegazzianum* biomes included) and (iii) Switzerland. The output maps of the world model were used as an environmental variables in the European model, and the output maps of the European model were in turn used as an environmental variables in for the Swiss model, after rescaling the 30 arc second resolution of the European model output maps to the 25m resolution of the Swiss environmental variables (Adde et al. 2023).

This was done for three reasons. 1) modelling the whole range, at least climatic, at which a species occurs is essential to avoid niche truncation, i.e., underestimation of the realized niche of the species by modelling only a fraction of the range of the environmental variables where the species occurs, 2) it allows for using the fine-resolution data available only at local (here Swiss) scale whilst also using coarser data at higher scales, and 3) it allows for selection of environmental variables most appropriate for predicting the distribution of *H. mantegazzianum* at each of the three extents.

The Maxent and Random Forest algorithms were used to fit the SDMs. These were chosen because they are both widely adopted algorithms (Gomes et al. 2018, Rathore and Sharma 2023) and because we found it appropriate to use one regression-based and one tree-based algorithm.

Maxent draws upon presence-only data, which is much more widely available than data on the absence of species (Guillera-Arroita et al. 2015; Fletcher et al. 2019). The version of the algorithm based on an infinitely weighted logistical regression (Phillips et al. 2017) was used here.

To avoid model overfitting of the Maxent models, we varied the mathematical function types (linear, quadratic, hinge, product, threshold) and the Regularization Multiplier (RM), a controller of model complexity: function types were varied linear, quadratic and hinge (LQH), 2) linear, quadratic and product (LQP), and 3) linear, quadratic product and hinge (LQPH). Threshold functions were never used because hinge functions have been shown to be a better replacement (Elith et al. 2011). In addition, the RM was varied from 1 to 5, giving us 15 combinations of settings.

We also used Random Forests, another widely used SDM algorithm. Tuning was done similar to the Maxent models, but the number of trees ("ntree" variable in randomForest function in randomForest package (Random Forests: https://www.stat.berkeley.edu/users/breiman/RandomForests/) was varied between 1000, 1500 and 2000 and the degree of random variation ("mtry" variable in randomForest function in randomForest package) was varied between 1 and 5. The selection of the optimal models was otherwise done in precisely the same way as for the Maxent models, but the performance metrics of the final models at all spatial extents were calculated using the out-of-bag (OOB) measures.

Background points selection

At each level of spatial analysis, the environmental variable raster layers were segmented into three strata using the following approach: first, we determined the range between the lowest and highest values within each layer. Next, we selected all the cells whose values fell within each one-third quantile of the dataset, creating three distinct strata within each layer. Each of these strata received a unique numerical identifier that did not repeat in any other layer. Subsequently, we generated a new raster layer in which these numeric codes were summed, resulting in a distinctive code for every strata combination. If the dataset was a factor (as is the case for the binary variables at European level), the cells containing “0” and “1” variables were replaced by a new unique code. Finally, we employed the 'sampleStratified' function from the raster R package to perform sampling across this newly created layer, thus generating stratified random points within the environmental space. The minimum number of points per stratum that yielded at least 10000 background points total were used. Since the model output at higher spatial extents was used as input for the model at lower spatial extent, the background points had to be resampled based on the occurrence point dataset used to generate the former. Therefore, each model at the European and Swiss spatial extents had a unique set of background points.

Performance metrics

We calculated the AUC (Hanley and McNeal 1984) and continuous Boyce index (Hirzel et al. 2006) performance metrics, as well as their sum (Steen at al. 2024b), 100x for each combination and selected the model settings (i.e., rm and function types for MaxEnt, ntree and mtry for RF) that had the highest metrics, splitting the occurrence data randomly between occurrence and validation bins each time. Before computing the sum, the AUC values were first recalculated to the gini AUC (2*AUC-1), in order to give it the same “weight” as the continuous Boyce index. The models with the highest performance metrics were projected on the matching environmental variables and the resulting rasters were summed in order to create the final product at each spatial extent.

If one or more of the models had the same settings, the output map of the model with those settings was used only once in the sum to create the final output maps. This is because the sum of the performance metrics was included in order to include models that had both high AUC and high Boyce, i.e., both high calibration and high discerning measures. If two models have

the same settings, then the sum has the same settings as the maximum AUC model, or the Maximum Boyce model, or both. Therefore, it would add no new information to the output to include it more than once and the final summed maps only included models with unique settings.

Variable importance

After selection of the models with the optimal performance metrics, the explanatory power of each of the predictors was determined for all the models. The varImportance() function of the fitMaxnet R package (Wilson 2024) was used for the Maxent models. For Random Forests, the necessary data is automatically saved when using the randomForest() function (Breiman et al. 2024).

Future projections

The CHELSA datasets that were projected by Karger et al. (2017) to the future time period of 2041-2060 using the SSP3-7.0 scenario were used in this study. This scenario was chosen because it occupies a middle ground between medium-high and very-high emissions. The scenario also assumes a fragmented world in which nations largely pursue their own interest and bars game-changing technological advancements (Karger et al. 2017). The world population model projections according to the SSP3-7.0 scenario for the year 2050 were used.

The future land use data were designed by Clarke lab (Esri Land Cover 2050) for the year 2050 (building on the data provided by the European Space Agency, which we used for present day land cover data) and were based on observed trends of historical land use change over the period 2010-2018. Whilst it does not particularly follow any RCP or SSP scenario, we believed that there were no conflicting assumptions made in any of the datasets. The future land use categories however did need to be harmonized with the current land use. The code used to accomplish this in the QGIS (QGIS Development Team 2009) raster calculator software is presented in Appendix 1.

There are however no future projections for most of the Swiss-scale model data, as the variables relating to topography were assumed to not change in the future. However, soil and edaphic variables will likely change. Therefore, we made the simplified assumption that the edaphic variables will not change. The future projections at the Swiss scale were therefore done by

inserting the outputs at the higher scales as an input variable, as part of our spatially nested approach.

Results

Fit of the models

Variable importance

The predictive power of all the included environmental variables differed slightly for each model fitted at each spatial extent but showed similar patterns overall. For the Maxent models fitted at world scale with non-thinned occurrence data, the annual temperature seasonality (bio15) had by far the highest predictive power, followed by the number of days in the growing season (gdd0). The same was true for the Maxent models fitted with occurrence data thinned with an MCP, but in the models fitted by data thinned by distance to artificial surfaces, the average temperature of the growth season (gst) was the second-strongest predictor (after bio15).

The RF models at world scale were also most powerfully predicted by bio15, with growth season temperature coming second. One of the two RF models fitted with MCP thinned data had Bio19 as the second-most powerful predictor and the other one had bio19 (mean monthly precipitation amount of the coldest quarter).

At Europe scale, by far the most powerful predictor was the world-scale model output. This held true across almost all models, except for Maxent and RF models fitted with data thinned by distance to artificial surface, in which it was distance to permanent snow and ice. Swampy or often flooded vegetation was the second-most powerful predictor for the other Maxent models. In RF models, distance to snow and ice ranked as the second most powerful predictor overall, though for one non-thinned dataset, population density was the second-most powerful predictor.

At Swiss scale, the Maxent models were most powerfully predicted by the slope, with the Europe-scale model variable coming second. For RF models, the most powerful predictor overall was the model fitted at European scale, though the slope was the strongest for RF models that used data thinned with an MCP, and the second-strongest predictor for the other models.

Model evaluation

Overfitting analysis

One Random Forest model at European scale (made with the MCP thinned occurrence dataset) was found to be overfitted (paired t-test: $p = 0.047$). However, as it was summed with other

MCP outputs at that scale and the distribution maps at Swiss scale are the main output of the study, this model was included in the results.

Performance metrics

Both Maxent and RF models showed good performance metrics, with average AUC at all scales always over 0.95 and average Boyce index always over 0.85. RF models had the highest average AUC, sometimes with values of around 0.999, especially for non-thinned data models, indicating an effect of high sample size. One RF model at European scale with non-thinned data (mtry = 2 and ntree = 2000) had only a Boyce index of 0.75, but this was a great outlier.

Present predictions

The final present-day outputs at Swiss scale are presented in Figure 3.

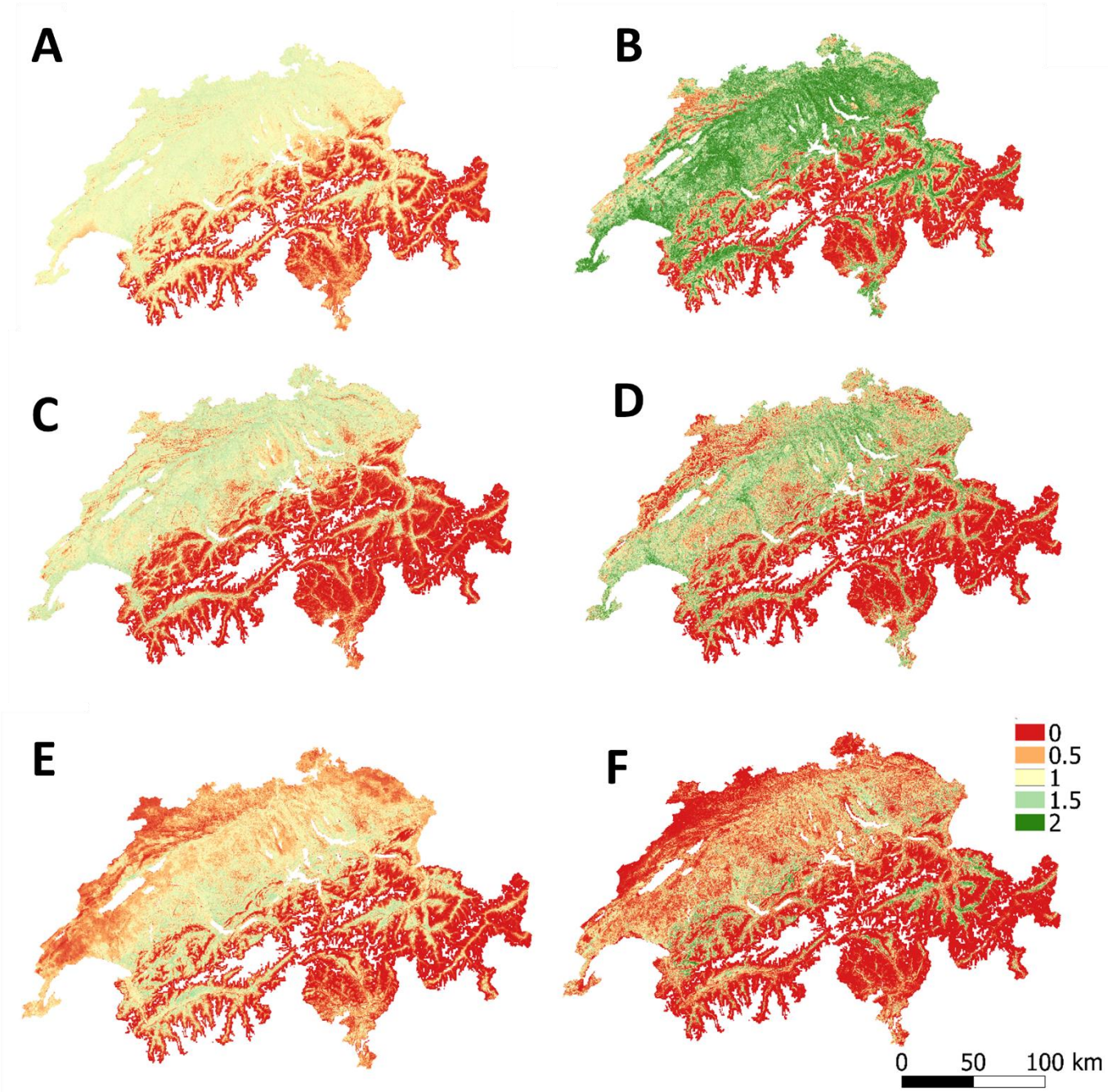

Figure 3: *Summed SDM outputs for* Heracleum mantegazzianum *for the present day, for both Maxent (A, C, E) and Random Forest (B, D, F) models, for all three occurrence datasets: non-thinned (A, B), thinned using a Minimum Convex Polygon (C, D) and thinned by distance to artificial surface and urban area (E, F). The scale bars denote habitat suitability and vary from 0-2 because there were always two models, with a habitat suitability range from 0 to 1.*

Future predictions

The final outputs at Swiss scale for the future are presented in Figure 4.

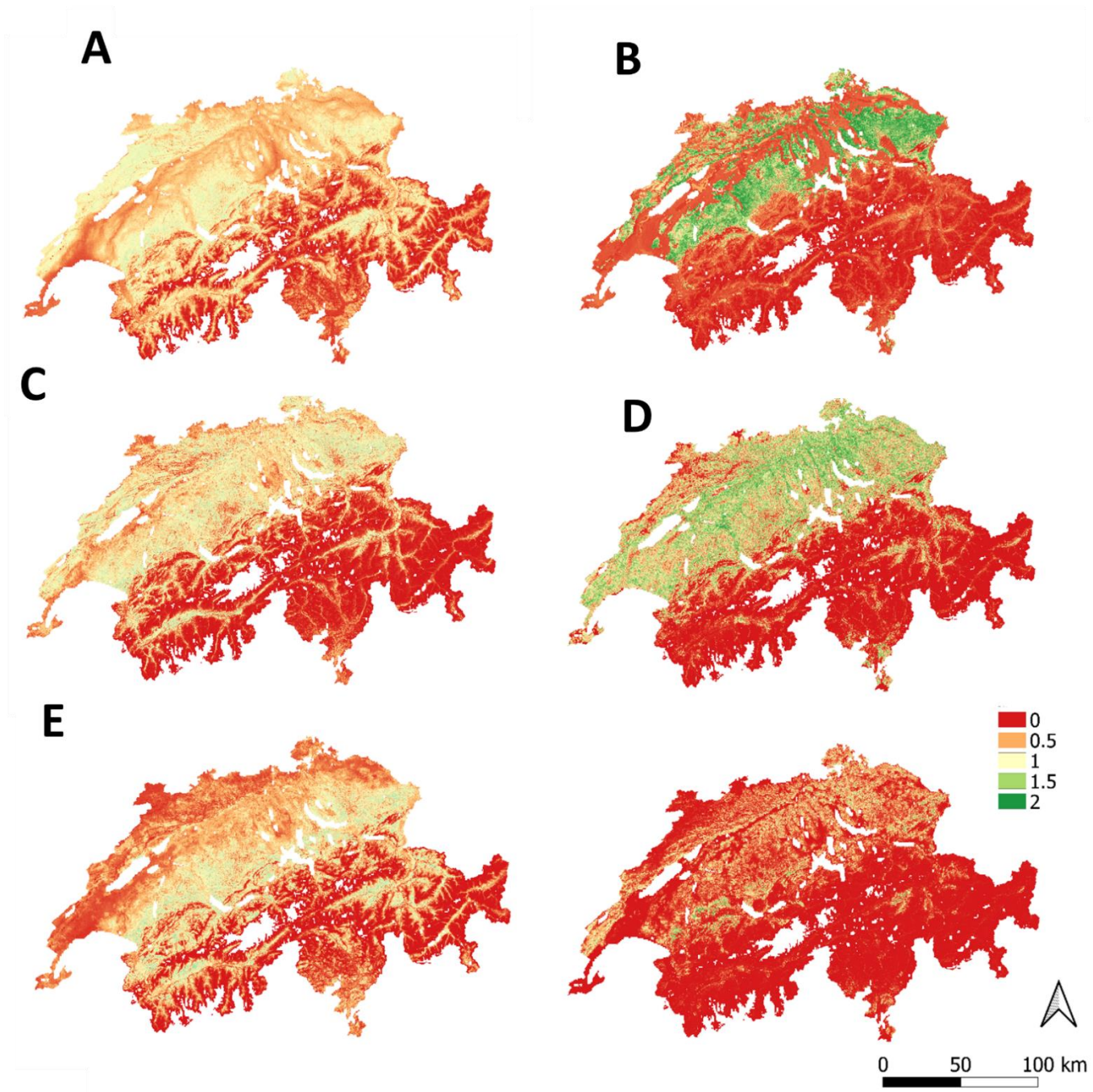

Figure 4: *Summed SDM outputs for* Heracleum mantegazzianum *for the year 2050, for both Maxent (A, C, E) and Random Forest (B, D, F) models, for all three occurrence datasets: non-thinned (A, B), thinned using a Minimum Convex Polygon (C, D) and thinned by distance to artificial surface and urban area (E, F). The scale bars denote habitat suitability and vary from 0-2 because there were always two models, with a habitat suitability range from 0 to 1.*

Discussion

Need for predicting and monitoring

Predicting and monitoring IAS is crucial for mitigating their negative impacts on biodiversity, ecosystem services, and human health (Shackleton et al. 2020). *Heracleum mantegazzianum* is a prime example of an IAS that poses significant threats to nature and economy and effectively managing it is therefore essential. Thankfully, it is possible to effectively manage the species (Nasadiuk and Mamchur 2024). These findings underscore the importance of continuous monitoring and adaptive management strategies to control the spread of invasive species like *H. mantegazzianum* and protect ecological integrity and ecosystem services. Here, we assessed how optimal predictive models could be built for this problematic IAS.

The Swiss Alps are refuges for biodiversity and provide invaluable ecosystem services (Ramel et al. 2020). With climate change, invasive alien plants are likely to shift their distributions to higher altitudes (Petitpierre et al. 2016), jeopardizing the biodiversity hotspots in and ecosystem services provided by the Swiss alps. For long, many mountains were thought to be invasion-free (Pauchard et al. 2009), but global warming is changing the story, and since *H. mantegazzian*um is a mountain species in its native range, its spread to higher altitudes is likely to happen where it invades (Cuddington et al. 2022). This is especially true considering that the species has demonstrated its ability to settle in high mountain areas when escaping from high-altitude botanical gardens (e.g. up to 2200 m; Müllerová 2024).

SDMs are increasingly recommended to be used in conservation studies (Guisan et al. 2013; Steen et al. 2024b), with even the International Union for the Conservation of Nature (IUCN) using them to estimate the species' distributions and to investigate how climate change will affect them (Cassini 2011). Studies have shown that SDMs can predict the potential spread of *H. mantegazzianum*, guiding targeted monitoring and management efforts (Shackleton et al. 2020). This study tested different settings to model the current and future distributions for *H. mantegazzianum* in Switzerland and look for the most optimal approach.

Current distributions

In the present-day models (Figure 3), the habitat suitability for *H. mantegazzianum* is lowest for the models thinned by proximity to artificial surface and urban areas. This approach assumes that citizen science data, such as most of the data from GBIF and Infoflora, are biased to

proximity to roads and also more easily accessible areas (Hughes et al. 2021, Díaz-Calafat et al. 2024 for insects). On the flipside, this approach truncates the niche of the species, by biasing the available data to areas further from infrastructure, which in Switzerland are more likely to be high-altitude areas, as the country is very mountainous and the vast majority of the population lives outside of the Alps (EDA, n.d.). It therefore stands to reason that the models that use data thinned by proximity to cities and roads model more rare environmental values, as mountain habitats have very variable climate and topography.

The fact that *H. mantegazzianum*'s distribution is most powerfully predicted by slope in Switzerland may reflect the fact that the species grows the most in the plains and lower lands (as seen in our occurrence data), presently not favoring the mountainous areas that cover much of Switzerland. This is supported by the fact that the plant indeed is more likely to occur at low slopes (see the response curves in Appendix 3).

Future distributions

Looking at the future predictions of habitat suitability in Switzerland (Figure 4), the habitat suitability decreases for all models, both Maxent and RF. However, in Maxent models, it does spread to higher up in mountains. This probably reflects the fact that *H. mantegazzianum* is a generalist species in its invaded range, meaning its distribution is mostly inside the most common climatic conditions found in Europe. This is supported by the fact that the model at world scale, fitted with climate variables, is the strongest predictor for the European models. The upward shift is likely representative of the future climatic shift, which would be driven by precipitation and the conditions of the growth season, as indicated by the most powerful predictors, the seasonality of precipitation, the length of the growth season and the temperature of the growth season.

The radical decrease of habitat suitability from present to future for the RF models fitted at Swiss scale with non-thinned occurrence data may be telling of model overfitting. This is more likely to occur with large datasets. The very sharp limit between suitable and unsuitable areas, particularly in the model fitted with non-thinned occurrence data, further supports this theory. In addition, our method of sampling background points stratified randomly in environmental space implies splitting each environmental variable up in three strata and sample equally from those three. The strongest predictor of the non-thinned data RF models is the summed model

output at European scale. For both RF Swiss-scale models with non-thinned data, the probability of occurrence of the species is almost entirely concentrated in the upper third of the variable values (see Appendix 3). It could therefore be that our stratified random in environmental space sampling allows the model to draw a very rigid line between suitable and unsuitable habitat, explaining the dramatic decrease in *H. mantegazzianum* habitat suitability in the future. The same holds true for the RF projections of the models fitted with data thinned by artificial surface and urban area.

It must also be noted that the decreases are more significant for RF models than for Maxent models, again perhaps indicating too strong fitting of data. However, the overfitting analyses of the RF models showed only one overfitted model on the European scale. Therefore, the models may perform very well, statistically, but put overly strict boundaries to the modelled niches. The more lenient predictions for Maxent models would seem to be more realistic for an invasive species, indicating that this algorithm is less inclined to put overly strict constraints on the modelled realized niche.

Implications for management

The usefulness of the output maps likely varies based on the exact spatial focus of conservation efforts. For instance, the models fitted with data from artificial surface and urban area do not match the present-day distribution of *H. mantegazzianum*. Therefore, these are likely not realistic. However, the Maxent models fitted with this dataset do demonstrate un upward shift of the species distribution in the future, and the output maps may therefore still be relevant when identifying areas to subject to scanning for the species in high-mountain habitats. Therefore, as all Maxent models fitted with all data display an upwards spread, they may all be useful for this purpose.

In general, the models fitted with MCP thinned data display the least dramatic shift, which is likely to match reality, due to *H. mantegazzianum*'s invasive and generalist nature. Still, the models fitted with non-thinned datasets display the highest suitability values in Switzerland and are therefore least likely to underestimate *H. mantegazzianum*'s potential distribution, which is important in the case of invasive alien species. However, to use this study's results to identify areas likely to be invaded, and therefore prioritize conservation efforts, it would be more prudent to use the slightly more restrictive MCP model outputs. Lastly, the output maps in this

study could be crossed with other relevant information, such as the spatial locations of biodiversity hotspots, the delivery of ecosystem services known to be impeded by *H. mantegazzianum* (such as recreational fishing) and locations where the health hazards posed by the species could be particularly harmful, such as in the proximity of schools.

Conclusions

The results presented here are a practical application of a tried-and-true method, i.e., the nested SDM approach that foils niche truncation and a new method, the use of environmentally stratified background points, to accurately predict the distributions of the invasive alien species *Heracleum mantegazzianum* Sommier & Levier. The background point sampling method has been demonstrated to increase model accuracy, though it possibly contributes to model overfitting. Whilst the model outputs differ between Random Forest and MaxEnt algorithms and by occurrence data thinning method, the models using occurrence data thinned by MCP likely represent the most reliable approximation of reality. Great conservation applications can be derived from this, when the 25m resolution Swiss maps are used to identify areas where *H. mantegazzianum* could cause significant health problems, such as densely populated areas and schools.

Acknowledgments

We thank La Sapienza University, the Swiss Confederation, the Institute of Earth Surface Dynamics of the University of Lausanne, and (in case of BS and AG) the Swiss national Valpar.CH project for providing research funding and support. We also thank the University of Lausanne for granting access to their High Performance Computing units.

References

Adde A, Rey P-L, Brun P, Külling N, Fopp F, Altermatt F, Broennimann O, Lehmann A, Petitpierre B, Zimmermann NE, Pellissier L, Guisan A (2023) N-SDM: a high-performance computing pipeline for Nested Species Distribution Modelling. Ecography 2023: e06540. https://doi.org/10.1111/ecog.06540

Anderson RP, Araújo MB, Guisan A, Lobo JM, Martínez-Meyer E, Peterson AT, Soberón JM (2020) Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography 12. https://doi.org/10.21425/F5FBG47839

Anderson RP, Araújo MB, Guisan A, Lobo JM, Martínez-Meyer E, Peterson AT, Soberón JM (2020) Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography 12. https://doi.org/10.21425/F5FBG47839

Aravind NA, Shaanker MU, Bhat H. N. P, Charles B, Shaanker R. U, Shah MA, G R (2022) Niche shift in invasive species: is it a case of “home away from home” or finding a “new home”? Biodiversity and Conservation 31: 2625–2638. https://doi.org/10.1007/s10531-022-02447-0

Global ESA CCI land cover classification map (1992-2015)- A Big Earth Data Platform for Three Poles. https://maps.elie.ucl.ac.be/CCI/viewer/

Bates OK, Ollier S, Bertelsmeier C (2020) Smaller climatic niche shifts in invasive than non-invasive alien ant species. Nature Communications 11: 5213. https://doi.org/10.1038/s41467-020-19031-1

Bhowmik PC, Chandran RS (2015) Biology, ecology, distribution and current status of Heracleum mantegazzianum Sommier and Levier. Journal of Crop and Weed 11: 1–17. https://cwss.in/Journal/Complete_jurnal/Vol.11-No.1__1.pdf

Bird TJ, Bates AE, Lefcheck JS, Hill NA, Thomson RJ, Edgar GJ, Stuart-Smith RD, Wotherspoon S, Krkosek M, Stuart-Smith JF, Pecl GT, Barrett N, Frusher S (2014) Statistical solutions for error and bias in global citizen science datasets. Biological Conservation 173: 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecological Modelling 157: 281–300. https://doi.org/10.1016/S0304-3800(02)00200-4

Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. Biology Letters 4: 585–589. https://doi.org/10.1098/rsbl.2008.0254

Caffrey JM (1999) Phenology and long-term control of Heracleum mantegazzianum. In: Caffrey J, Barrett PRF, Ferreira MT, Moreira IS, Murphy KJ, Wade PM (Eds), Biology, Ecology and Management of Aquatic Plants. Springer Netherlands, Dordrecht, 223–228. https://doi.org/10.1007/978-94-017-0922-4_31

Callaghan CT, Rowley JJL, Cornwell WK, Poore AGB, Major RE (2019) Improving big citizen science data: Moving beyond haphazard sampling. PLOS Biology 17: e3000357. https://doi.org/10.1371/journal.pbio.3000357

Cassini MH (2011) Ranking threats using species distribution models in the iucn red list assessment process. Biodiversity and Conservation 20: 3689–3692. https://doi.org/10.1007/s10531-011-0126-9

Celesti-Grapow L, Brundu G, Citterio S, Galasso G, Gentili R, Alessandrini A, Argenti C, et al. (submitted) Distribution of the invasive Heracleum mantegazzianum in Italy. Bioinvasion Records.

Chevalier M, Broennimann O, Cornuault J, Guisan A (2021) Data integration methods to account for spatial niche truncation effects in regional projections of species distribution. Ecological Applications: A Publication of the Ecological Society of America 31: e02427. https://doi.org/10.1002/eap.2427

Cuddington K, Sobek-Swant S, Drake J, Lee W, Brook M (2022) Risks of giant hogweed (Heracleum mantegazzianum) range increase in North America. Biological Invasions 24: 299–314. https://doi.org/10.1007/s10530-021-02645-x

Díaz-Calafat J, Jaume-Ramis S, Soacha K, Álvarez A, Piera J (2024) Revealing biases in insect observations: A comparative analysis between academic and citizen science data. PLOS ONE 19: e0305757. https://doi.org/10.1371/journal.pone.0305757

Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH (2021) Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. BioScience 71: 1179–1188. https://doi.org/10.1093/biosci/biab093

Di Cola V, Broennimann O, Petitpierre B, Breiner FT, D'Amen M, Randin C, Engler R, Pottier J, Pio D, Dubuis A, Pellissier L, Mateo RG, Hordijk W, Salamin N, Guisan A (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. Ecography 40: 774–787. https://doi.org/10.1111/ecog.02671

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, Mcclean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27–46.

Esri Land Cover 2050. https://www.arcgis.com/home/item.html?id=b5ee7191eda1425fa18c26532683896d

Earth Science Data Systems N (2024) Global 1-km Downscaled Population Base Year and Projection Grids Based on the Shared Socioeconomic Pathways, Revision 01 | NASA Earthdata. Available from: https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-pd-sspbsyr-1km-1.01 (December 9, 2024).

EASIN - European Alien Species Information Network. https://easin.jrc.ec.europa.eu/spexplorer/search/searchpaged

European Commission - Invasive alien species. https://environment.ec.europa.eu/topics/nature-and-biodiversity/invasive-alien-species_en

Fernández M, Hamilton H (2015) Ecological Niche Transferability Using Invasive Species as a Case Study. PLOS ONE 10: e0119891. https://doi.org/10.1371/journal.pone.0119891

Fletcher Jr. RJ, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM (2019) A practical guide for combining data to model species distributions. Ecology 100: e02710. https://doi.org/10.1002/ecy.2710

FOAG (2012). *Carte des aptitudes des sols de la Suisse*. https://www.blw.admin.ch/fr/carte-des-aptitudes-des-sols

Franklin J (2010) Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511810602

Gallien L, Douzet R, Pratte S, Zimmermann NE, Thuiller W (2012) Invasive species distribution models – how violating the equilibrium assumption can create new insights. Global Ecology and Biogeography 21: 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

GBIF.org (09 October 2021) GBIF Occurrence Download
Geography – Facts and Figures. https://www.eda.admin.ch/aboutswitzerland/en/home/umwelt/geografie/geografie---fakten-und-zahlen.html

Gomes VHF, IJff SD, Raes N, Amaral IL, Salomão RP, de Souza Coelho L, de Almeida Matos FD, Castilho CV, de Andrade Lima Filho D, López DC, Guevara JE, Magnusson WE, Phillips OL, Wittmann F, de Jesus Veiga Carim M, Martins MP, Irume MV, Sabatier D, Molino J-F, Bánki OS, da Silva Guimarães JR, Pitman NCA, Piedade MTF, Mendoza AM, Luize BG, Venticinque EM, de Leão Novo EMM, Vargas PN, Silva TSF, Manzatto AG, Terborgh J, Reis NFC, Montero JC, Casula KR, Marimon BS, Marimon B-H, Coronado ENH, Feldpausch TR, Duque A, Zartman CE, Arboleda NC, Killeen TJ, Mostacedo B, Vasquez R, Schöngart J, Assis RL, Medeiros MB, Simon MF, Andrade A, Laurance WF, Camargo JL, Demarchi LO, Laurance SGW, de Sousa Farias E, Nascimento HEM, Revilla JDC, Quaresma A, Costa FRC, Vieira ICG, Cintra BBL, Castellanos H, Brienen R, Stevenson PR, Feitosa Y, Duivenvoorden JF, Aymard C. GA, Mogollón HF, Targhetta N, Comiskey JA, Vicentini A, Lopes A, Damasco G, Dávila N, García-Villacorta R, Levis C, Schietti J, Souza P, Emilio T, Alonso A, Neill D, Dallmeier F, Ferreira LV, Araujo-Murakami A, Praia D, do Amaral DD, Carvalho FA, de Souza FC, Feeley K, Arroyo L, Pansonato MP, Gribel R, Villa B, Licona JC, Fine PVA, Cerón C, Baraloto C, Jimenez EM, Stropp J, Engel J, Silveira M, Mora MCP, Petronelli P, Maas P, Thomas-Caesar R, Henkel TW, Daly D, Paredes MR, Baker TR, Fuentes A, Peres CA, Chave J, Pena JLM, Dexter KG, Silman MR, Jørgensen PM, Pennington T, Di Fiore A, Valverde FC, Phillips JF, Rivas-Torres G, von Hildebrand P, van Andel TR, Ruschel AR, Prieto A, Rudas A, Hoffman B, Vela CIA, Barbosa EM, Zent EL, Gonzales GPG, Doza HPD, de Andrade Miranda IP, Guillaumet J-L, Pinto LFM, de Matos Bonates LC, Silva N, Gómez RZ, Zent S, Gonzales T, Vos VA, Malhi Y, Oliveira AA, Cano A, Albuquerque BW, Vriesendorp C, Correa DF, Torre EV, van der Heijden G, Ramirez-Angulo H, Ramos JF, Young KR, Rocha M, Nascimento MT, Medina MNU, Tirado M, Wang O, Sierra R, Torres-Lezama A, Mendoza C, Ferreira C, Baider C, Villarroel D, Balslev H, Mesones I, Giraldo LEU, Casas LF, Reategui MAA, Linares-

Palomino R, Zagt R, Cárdenas S, Farfan-Rios W, Sampaio AF, Pauletto D, Sandoval EHV, Arevalo FR, Huamantupa-Chuquimaco I, Garcia-Cabrera K, Hernandez L, Gamarra LV, Alexiades MN, Pansini S, Cuenca WP, Milliken W, Ricardo J, Lopez-Gonzalez G, Pos E, ter Steege H (2018) Species Distribution Modelling: Contrasting presence-only models with plot abundance data. Scientific Reports 8: 1–12. https://doi.org/10.1038/s41598-017-18927-1

Grêt-Regamey A, Kytzia S (2007) Integrating the valuation of ecosystem services into the Input–Output economics of an Alpine region. Ecological Economics 63: 786–798. https://doi.org/10.1016/j.ecolecon.2007.02.026

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA (2015) Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography 24: 276–292. https://doi.org/10.1111/geb.12268

Guisan A, Thuiller W, Zimmermann NE (2017) Habitat Suitability and Distribution Models: With Applications in R. https://doi.org/10.1017/9781139028271

Guisan A, Chevalier M, Adde A, Zarzo-Arias A, Goicolea T, Broennimann O, Petitpierre B, Scherrer D, Rey P-L, Collart F, Riva F, Steen B, Mateo RG Spatially nested species distribution models (N-SDM): An effective tool to overcome niche truncation for more robust inference and projections. Journal of Ecology n/a. https://doi.org/10.1111/1365-2745.70063

Haider S, Lembrechts JJ, McDougall K, Pauchard A, Alexander JM, Barros A, Cavieres LA, Rashid I, Rew LJ, Aleksanyan A, Arévalo JR, Aschero V, Chisholm C, Clark VR, Clavel J, Daehler C, Dar PA, Dietz H, Dimarco RD, Edwards P, Essl F, Fuentes-Lillo E, Guisan A, Gwate O, Hargreaves AL, Jakobs G, Jiménez A, Kardol P, Kueffer C, Larson C, Lenoir J, Lenzner B, Padrón Mederos MA, Mihoc M, Milbau A, Morgan JW, Müllerová J, Naylor BJ, Nijs I, Nuñez MA, Otto R, Preuk N, Ratier Backes A, Reshi ZA, Rumpf SB, Sandoya V, Schroder M, Speziale KL, Urbach D, Valencia G, Vandvik V, Vitková M, Vorstenbosch T, Walker TWN, Walsh N, Wright G, Zong S, Seipel T (2022) Think globally, measure locally: The MIREN standardized protocol for monitoring plant species distributions along elevation gradients. Ecology and Evolution 12: e8590. https://doi.org/10.1002/ece3.8590

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling 199: 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

Hughes AC, Orr MC, Ma K, Costello MJ, Waller J, Provoost P, Yang Q, Zhu C, Qiao H (2021) Sampling biases shape our view of the natural world. Ecography 44: 1259–1269. https://doi.org/10.1111/ecog.05926

Infoflora Lists. https://www.infoflora.ch/en/neophytes/lists.html

IPBES (2023) Summary for Policymakers of the Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on

Biodiversity and Ecosystem Services. In: Roy HE, Pauchard A, Stoett P, Renard Truong T, Bacher S, Galil BS, Hulme PE, Ikeda T, Sankaran KV, McGeoch MA, Meyerson LA, Nuñez MA, Ordonez A, Rahlao SJ, Schwindt E, Seebens H, Sheppard AW, Vandvik V (Eds), IPBES Secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.7430692

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M (2017) Climatologies at high resolution for the earth's land surface areas. Scientific Data 4: 170122. https://doi.org/10.1038/sdata.2017.122

Kass JM, Smith AB, Warren DL, Vignali S, Schmitt S, Aiello-Lammens ME, Arlé E, Márcia Barbosa A, Broennimann O, Cobos ME, Guéguen M, Guisan A, Merow C, Naimi B, Nobis MP, Ondo I, Osorio-Olvera L, Owens HL, Pinilla-Buitrago GE, Sánchez-Tapia A, Thuiller W, Valavi R, Velazco SJE, Zizka A, Zurell D (2025) Achieving higher standards in species distribution modeling by leveraging the diversity of available software. Ecography 2025: e07346. https://doi.org/10.1111/ecog.07346

Külling N, Adde A, Fopp F, Schweiger AK, Broennimann O, Rey P-L, Giuliani G, Goicolea T, Petitpierre B, Zimmermann NE, Pellissier L, Altermatt F, Lehmann A, Guisan A (2024) SWECO25: a cross-thematic raster database for ecological research in Switzerland. Scientific Data 11: 21. https://doi.org/10.1038/s41597-023-02899-1

Lake TA, Briscoe Runquist RD, Moeller DA (2020) Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections. Diversity and Distributions 26: 1767–1779. https://doi.org/10.1111/ddi.13161

McGill BJ (2010) Matters of Scale. Science 328: 575–576. https://doi.org/10.1126/science.1188528

Mohr CO (1947) Table of Equivalent Populations of North American Small Mammals. American Midland Naturalist 37: 223. https://doi.org/10.2307/2421652

Müllerová J (2024) Heracleum mantegazzianum (giant hogweed). CABI Compendium CABI Compendium: 26911. https://doi.org/10.1079/cabicompendium.26911

Nasadiuk P, Mamchur Z (2024) Review of modern approaches to the management of invasive species of the genus Heracleum in Europe. Visnyk of Lviv University. Biological series: 67–89. https://doi.org/10.30970/vlubs.2024.92.06

Ni J, Wang GH, Bai YF, Li XZ (2007) Scale-dependent relationships between plant diversity and above-ground biomass in temperate grasslands, south-eastern Mongolia. Journal of Arid Environments 68: 132–142. https://doi.org/10.1016/j.jaridenv.2006.05.003

Pauchard A, Kueffer C, Dietz H, Daehler CC, Alexander J, Edwards PJ, Arévalo JR, Cavieres LA, Guisan A, Haider S, Jakobs G, McDougall K, Millar CI, Naylor BJ, Parks CG, Rew LJ, Seipel T (2009) Ain't no mountain high enough: plant invasions reaching new elevations. Frontiers in Ecology and the Environment 7: 479–486. https://doi.org/10.1890/080072

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12: 361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27: 285–298. https://doi.org/10.1111/j.0906-7590.2004.03740.x

Pergl J, Müllerová J, Perglová I, Herben T, Pyšek P (2011) The role of long-distance seed dispersal in the local population dynamics of an invasive plant species. Diversity and Distributions 17: 725–738. https://doi.org/10.1111/j.1472-4642.2011.00771.x

Petitpierre B, McDougall K, Seipel T, Broennimann O, Guisan A, Kueffer C (2016) Will climate change increase the risk of plant invasions into mountains? Ecological Applications 26: 530–544. https://doi.org/10.1890/14-1871

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography 40: 887–893. https://doi.org/10.1111/ecog.03049

Pysek P, Cock MJW, Nentwig W, Ravn HP (2007) Ecology and Management of Giant Hogweed. Illustrated edition. CABI, Wallingford, 352 pp.

Spatial without Compromise · QGIS Web Site. Available from: https://qgis.org/ (December 9, 2024).

Ramel C, Rey P-L, Fernandes R, Vincent C, Cardoso AR, Broennimann O, Pellissier L, Pradervand J-N, Ursenbacher S, Schmidt BR, Guisan A (2020) Integrating ecosystem services within spatial biodiversity conservation prioritization in the Alps. Ecosystem Services 45: 101186. https://doi.org/10.1016/j.ecoser.2020.101186

Rathore MK, Sharma LK (2023) Efficacy of species distribution models (SDMs) for ecological realms to ascertain biological conservation and practices. Biodiversity and Conservation 32: 3053–3087. https://doi.org/10.1007/s10531-023-02648-1

Random Forests. Available from: https://www.stat.berkeley.edu/users/breiman/RandomForests/ (December 9, 2024).

Breiman L, Cutler A, Liaw A, Wiener M (2024) randomForest: Breiman and Cutlers Random Forests for Classification and Regression. Available from: https://cran.r-project.org/web/packages/randomForest/index.html (January 17, 2025).

a. Steen B, Adde A, Schlaepfer MA, Guisan A, Maiorano L (2024) Distributions of non-native and native plants are not determined by the same environmental factors. Ecological Solutions and Evidence 5: e12374. https://doi.org/10.1002/2688-8319.12374

b. Steen B, Broennimann O, Maiorano L, Guisan A (2024) How sensitive are species distribution models to different background point selection strategies? A test with species at various equilibrium levels. Ecological Modelling 493: 110754. https://doi.org/10.1016/j.ecolmodel.2024.110754

Steen VA, Tingley MW, Paton PWC, Elphick CS (2021) Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen

science data. Methods in Ecology and Evolution 12: 216–226. https://doi.org/10.1111/2041-210X.13525

Thiele J, Otte A, Eckstein RL (2007) Ecological needs, habitat preferences and plant communities invaded by Heracleum mantegazzianum. In: Ecology and management of giant hogweed (Heracleum mantegazzianum). CABI Books, 126–143. https://doi.org/10.1079/9781845932060.0126

Thiele J, Otte A (2008) Invasion patterns of *Heracleum mantegazzianum* in Germany on the regional and landscape scales. Journal for Nature Conservation 16: 61–71. https://doi.org/10.1016/j.jnc.2007.08.002

Thiele J, Schuckert U, Otte A (2008) Cultural landscapes of Germany are patch-corridor-matrix mosaics for an invasive megaforb. Landscape Ecology 23: 453–465. https://doi.org/10.1007/s10980-008-9202-2

Tulloch AIT, Sutcliffe P, Naujokaitis-Lewis I, Tingley R, Brotons L, Ferraz KMPMB, Possingham H, Guisan A, Rhodes JR (2016) Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. Biological Conservation 199: 157–171. https://doi.org/10.1016/j.biocon.2016.04.023

Environment UN (2023) Invasive Alien Species Report | UNEP - UN Environment Programme. Available from: https://www.unep.org/resources/report/invasive-alien-species-report (December 9, 2024).

Vicente JR, Gonçalves J, Honrado JP, Randin CF, Pottier J, Broennimann O, Lomba A, Guisan A (2014) A framework for assessing the scale of influence of environmental factors on ecological patterns. Ecological Complexity 20: 151–156. https://doi.org/10.1016/j.ecocom.2014.10.005

Wilson PD (2024) peterbat1/fitMaxnet. Available from: https://github.com/peterbat1/fitMaxnet (January 16, 2025).