AGILE: GIScience Series, 6, 37, 2025. https://doi.org/10.5194/agile-giss-6-37-2025
Proceedings of the 28th AGILE Conference on Geographic Information Science, 10–13 June 2025.
Eds.: Auriol Degbelo, Serena Coetzee, Carsten Keßler, Monika Sester, Sabine Timpf, Lars Bernard
This contribution underwent peer review based on a full paper submission.

Evaluating the role of human and environmental factors causing the distribution of invasive plant species in the cantons of Vaud and Neuchâtel in Switzerland

Maryam Lotfian [1], Simon Walther[2], Magali Egger[3], Olivier D'Ancona[3], Andres Perez-Uribe [2], and Jens Ingensand [1]

[1]Institute INSIT, School of Engineering and Management Vaud, University of Applied Sciences and Arts Western Switzerland, Yverdon-les-Bains, Switzerland
[2]Institute IICT, School of Engineering and Management Vaud, University of Applied Sciences and Arts Western Switzerland, Yverdon-les-Bains, Switzerland
[3]HES-SO University of Applied Sciences and Arts Western Switzerland, Lausanne, Switzerland

Abstract. Invasive alien plant species are an increasing concern in many countries due to their negative impacts on local ecosystems, human health, infrastructure, and agriculture, to name a few. In Switzerland, substantial financial resources are allocated each year to combat the spread and eradicate these invasive species. Understanding their spatial distribution through species distribution modeling is crucial for improving management interventions. This study aims to examine the role of environmental and human factors in predicting the distribution of three invasive plant species (Prunus laurocerasus, Buddleja davidii, and Robinia pseudoacacia) in the Cantons of Vaud and Neuchâtel in Western Switzerland. A random forest algorithm is trained, and the resulting model is used to assess the relative importance of various environmental and human factors in predicting species distribution. The results highlight that while environmental features play a significant role in generating distribution maps, incorporating human activity patterns, such as proximity to built areas, railways, and roads, greatly enhances prediction accuracy and leads to more robust models.

Submission Type. model, analysis, dataset, case study

BoK Concepts. EO for biodiversity and ecosystems (TA12-2), Climate, Environment, and Biodiversity (SA3-4-1-3), Use of geospatial information in environmental issues (GS3-4)

Keywords. Invasive Plants, Biodiversity, Citizen Science, Species Distribution Modeling, Random Forest

1 Introduction

The proliferation of invasive alien species is a significant concern worldwide (Andersen et al., 2004; Stohlgren and Schnase, 2005). These species can disrupt native ecosystems, leading to biodiversity loss and substantial economic costs (Linders et al., 2019). A 2023 report by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) highlighted that invasive alien species have been a major factor in 60 percent of global animal and plant extinctions, and the sole driver in 16 per cent of these cases (Roy et al., 2024). The primary driver behind the introduction of invasive species is the movement of goods and people, which leads to their intentional or unintentional introduction into new environments (Keller et al., 2011).

Invasive alien plant species is becoming a growing concern in Europe (European Commission DG Environment, 2024). In Switzerland, extensive efforts are underway to control their spread, requiring considerable human and financial resources (Federal Office for the Environment (FOEN), 2022). Beyond their ecological impact, invasive plant species pose threats to infrastructure such as roads and railways, disrupt agricultural productivity, and, in some cases, even have adverse health effects on humans (Kumar Rai and Singh, 2020). A deeper understanding of their distribution patterns can facilitate more efficient management strategies and targeted interventions.

Predicting the geographical distribution of invasive plant species is a key component of their control and potential eradication (Jarnevich et al., 2023). Identifying the environmental and human-related factors that drive their spread can lead to more effective mitigation and management measures.

This study evaluates the distribution of invasive plant species in Switzerland, with a particular focus on three high-priority species for eradication in western Switzerland: Prunus laurocerasus, Buddleja davidii, and Robinia pseudoacacia. The research is conducted in the cantons of Vaud and Neuchâtel. The primary objective is to utilize environmental and climatic variables as well as information on human settlements to model the distribution of the mentioned invasive plant species and address the following research question:

Can the integration of environmental factors and human activity patterns enhance predictions of invasive species spread and inform more effective management strategies?

The structure of this article is as follows: The next section discusses the role of species distribution modeling in policymaking. This is followed by a detailed methodology section covering dataset description, data processing, exploratory analysis, and modeling approach. The results of the modeling are then presented, followed by a discussion of key findings and concluding remarks.

2 Species Distribution Modeling for Policy Management

Species Distribution Modeling (SDM) has been widely used for many years to predict the most probable areas for species presence based on a combination of environmental and physical factors (Elith and Leathwick, 2009; Miller, 2010). SDM plays a crucial role in biodiversity conservation efforts, helping policymakers in identifying priority areas for intervention as well as biodiversity loss hotspots (McSHEA, 2014).

For invasive plant species, SDM is particularly effective in supporting eradication efforts by identifying areas with a high probability of species occurrence and evaluating the factors contributing to their proliferation (Jarnevich et al., 2023). Additionally, SDM can be useful in monitoring eradication success by assessing whether previously targeted areas remain free of the invasive species or if recolonization has occurred (Cho et al., 2022).

There are various SDM methodologies, among which MaxEnt(Elith et al., 2010) is one of the most known for ecological modeling. MaxEnt relies on presence data (locations where the species has been observed) and and contrasts them with background data sampled from across the study area to estimate the relative environmental suitability of different locations. In contrast, many machine learning models require both presence and absence data for more comprehensive modeling(Elith et al., 2020; Lotfian et al., 2022). This study employs the Random Forest algorithm to generate distribution models for three selected invasive plant species. The reason for the choice of Random Forest is that it is a robust ensemble learning method that handles non-linear relationships, accommodates high-dimensional datasets, and automatically captures interactions among predictor variables (Simon et al., 2023). Additionally, it provides measures of variable importance, which can support interpretation and inform management decisions (Fox et al., 2017).

The following section presents the dataset and the methodologies employed for modeling.

3 Methodology

3.1 Dataset

The dataset used in this study consists of records on invasive plant species and various environmental and climatic variables. The key components of the dataset are detailed below:

Invasive plant species: This study focuses on two cantons in western Switzerland—Vaud and Neuchâtel—located in the French-speaking region. Data on invasive alien plant species for these cantons were obtained from the National Data and Information Center on Swiss Flora, known as Info Flora (2025) (see Figure 1).

Info Flora utilizes the citizen science platform *InvasiveApp* to collect observations on invasive plant species. While data quality and participation inequality are common challenges in participatory science projects (Moradi et al., 2021; Lotfian et al., 2023), citizen-generated data often represents the best available resource, particularly for ecological research (Callaghan et al., 2024). Info Flora's dataset is no exception to these challenges, including spatial biases, as seen in Figure 1. However, despite these limitations, it remains one of the most comprehensive and valuable sources of plant species data in Switzerland, thanks to the dedicated contributions of volunteers.

The dataset from Info Flora includes 39 features that can be categorized into five groups: observation details, taxonomic information, geographical information, management and invasive status, and abundance/phenology.

Although absence data are available in the dataset, they are extremely limited. For instance, in the Info Flora dataset, presence points for *Prunus laurocerasus*, *Buddleja davidii*, and *Robinia pseudoacacia* are in total 2,829, 5,390, and 3,997, respectively. However, only a single absence record exists for *Buddleja davidii*. Therefore, we generated pseudo-absence (background) data to enhance model performance.

To create the pseudo-absence/background dataset, we randomly generated absence points in locations where presence points were not recorded. To reduce spatial bias, we applied a 1 km buffer around each presence point and randomly generated an equal number of pseudo-absence points for each species outside these buffered areas.

Land use: To assess the impact of human infrastructure on the spread of invasive plants, we introduced a land-use feature that quantifies proximity to various infrastructure

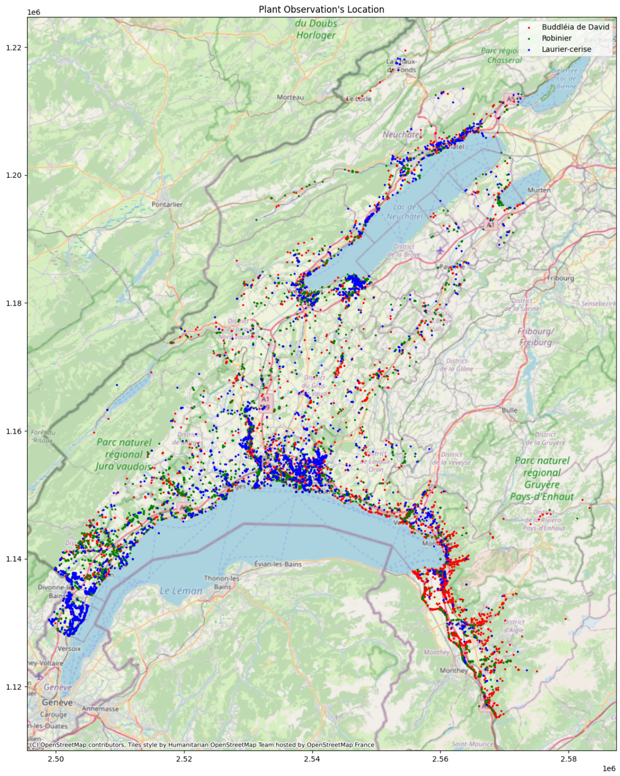

Figure 1. The observations points from Info Flora for the three selected invasive plant species within two cantons of Vaud and Neuchâtel in Western Switzerland. Prunus laurocerasus (blue), Buddleja davidii (red), and Robinia pseudoacacia (green)

elements. OpenStreetMap (OSM) data were accessed via the *Overpass API*, using the *OSMnx* Python library. We extracted features within a 250 m buffer around each observation, including:

- Building areas (average and total)
- Distance to the nearest buildings
- Distance to highways, railways, and waterways

Meteorological data: To account for local climate conditions, we obtained yearly normals of temperature, precipitation, and sunshine from the Federal Office of Meteorology and Climatology (MeteoSwiss). This data, provided in *NetCDF* format, and has a spatial resolution of 1 km.

NDVI (Normalized Difference Vegetation Index): To capture vegetation health and density around observation points, we used Landsat 8 satellite imagery to compute NDVI at a 30 m resolution. For each observation point, we calculated the mean and standard deviation of NDVI values within 30 m, 250 m, and 500 m buffers. To ensure accurate representation of peak vegetation activity, we focused on NDVI values from the summer months.

Altitude: Finally, we obtained altitude data from Federal Office of Topography known as swisstopo (2025) at a 2 m resolution to compute the average elevation and slope surrounding each observation point.

3.2 Exploratory analysis and data pre-processing

Before training the model, an exploratory analysis was conducted to visualize the behavior of the three selected plant species. This included using box plots to visualize the distribution of species based on average annual climatic conditions, average elevation, and proximity to different land-use types. These visualizations provided an initial understanding of the spatial distribution of data points within the study area.

Next, the data was prepared for the machine learning model. To account for spatial autocorrelation, we avoided randomly splitting the dataset into train, test, and validation sets. Instead, we first applied the k-means clustering algorithm to group nearby observations. Once clustering was complete, the data was divided into training (70%), testing (15%), and validation (15%) sets (See Figure 2).

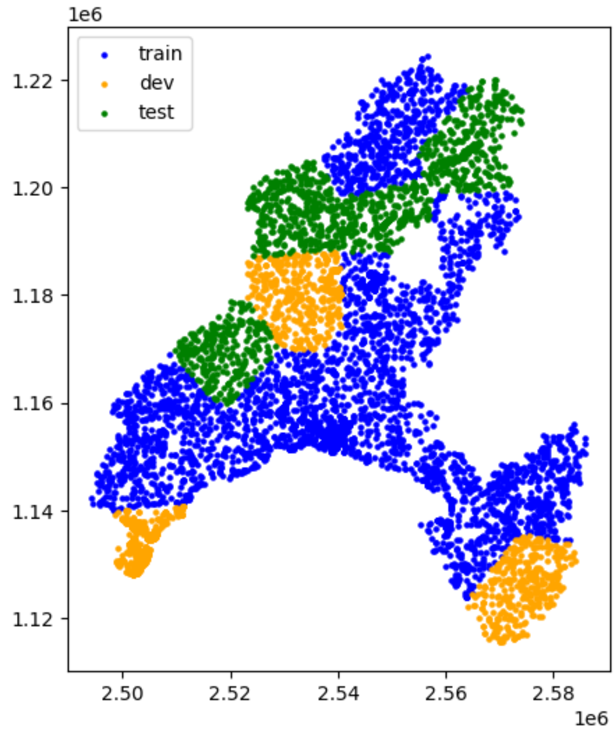

Figure 2. The observations were split into clusters using K-means clustering. This similarity approach ensures that the training, validation, and test sets are spatially distinct, reducing spatial autocorrelation biases.

3.3 Model training

As mentioned earlier, this study employs the Random Forest algorithm to predict the distribution of the three invasive plant species. Random Forest has been widely recognized in the literature as an effective method for tabular data analysis (Grinsztajn et al., 2022). We chose Random Forest because they efficiently handle both numerical and categorical data, perform well even with default settings, and provide insights into feature importance, an essential aspect of ecological studies. Additionally, this ensemble-based approach is robust to outliers and captures complex interactions among predictors, making it particularly useful in ecological modeling, where multiple factors influence plant presence.

For implementation, we used the *scikit-learn* library, keeping most parameters at their default values since they yielded satisfactory results in preliminary tests. Specifically, we set the number of estimators (trees) to 100, used the Gini criterion for determining splits, and retained the default maximum tree depth (None), allowing trees to grow until all leaves are pure or contain fewer than the minimum split samples. While further hyperparameter tuning may improve performance, our primary objective was to demonstrate the feasibility of using an enriched dataset to enhance predictions of invasive plant presence.

3.4 Data and Software Availability

The code and the dataset used in this research can be found here at Zenodo.

4 Results

4.1 Explanatory analysis:

A preliminary analysis was conducted to assess the distribution of observation points across different land use types, climatic conditions, and altitudinal ranges.

The box plots in Figure 3 illustrate that the distributions of the three climatic variables for the three species are relatively similar. The data suggests that these species thrive better in regions with higher mean yearly temperatures, as reflected by the medians near 11°C. While they are adaptable as well to cooler conditions, their primary distribution aligns with warmer climates, which may enhance their invasive potential in such areas. Additionally, the box plots show that, beyond temperature, broader adaptability to precipitation and sunshine duration—particularly for Buddleja davidii—suggests that temperature alone does not dictate their invasive success. Instead, their ability to persist across diverse environmental conditions, coupled with moderate climatic conditions, likely plays a significant role.

For land use analysis, the proximity to different land use types indicates that the three species are predominantly observed in forests, grasslands, and residential areas. Prunus laurocerasus is particularly common in residential zones (Figure 4), where it is frequently used as a hedge plant in private gardens to provide privacy.

As expected, since the species are predominantly observed in urban areas, their presence points are concentrated at lower altitudes. Figure 5 illustrates that Prunus laurocerasus is primarily distributed around 400 meters in altitude, which aligns with its preference for residential environments.

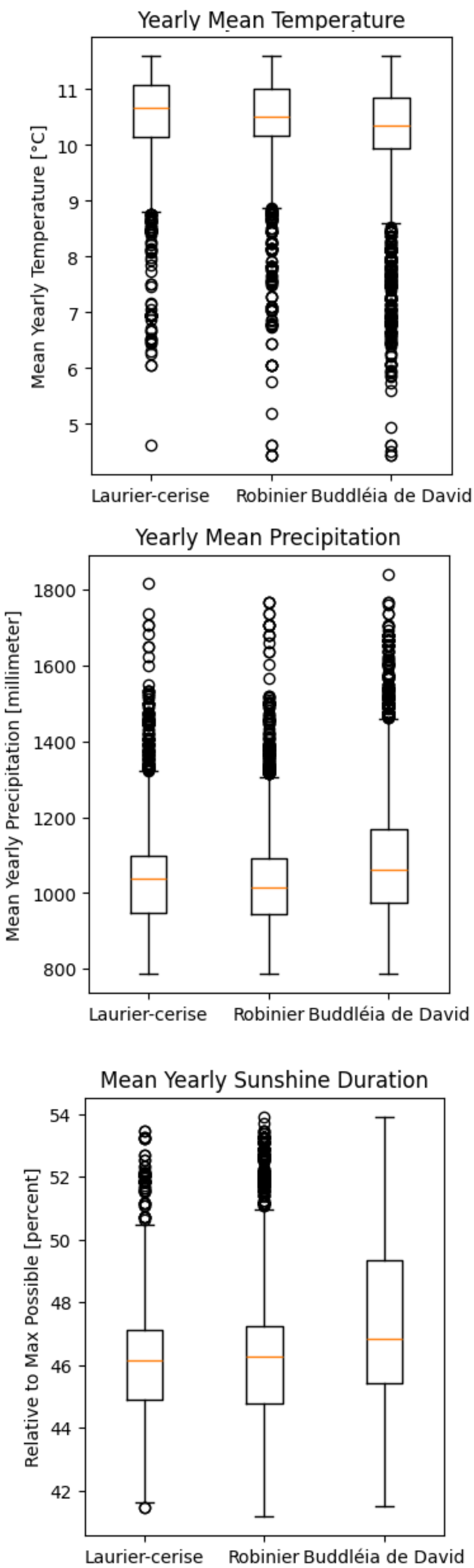

Figure 3. Comparison of meteorological variations (yearly mean temperature, precipitation, and sunshine duration) for the three selected invasive plants and across canton Vaud and Neuchâtel, Switzerland.

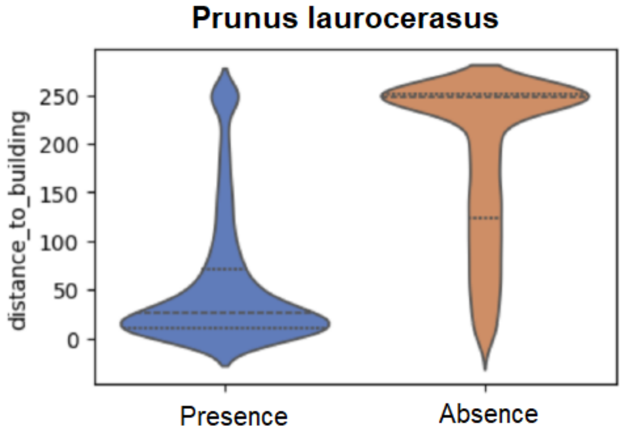

Figure 4. Distance to building for presence versus absence points of Prunus laurocerasus species. Presence points are mostly distributed at close vicinity of the buildings.

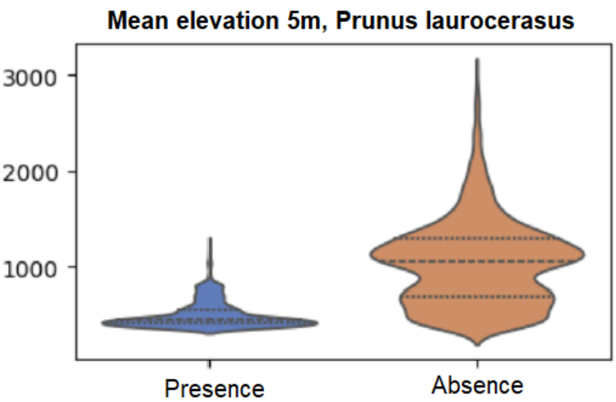

Figure 5. The distribution of presence points of Prunus laurocerasus versus the absence points, across various altitudes, with an average altitude calculated within a 5-meter radius around each point.

4.2 Model evaluation and feature importance

Model Evaluation: The model was evaluated on a held-out development set to assess its performance. The confusion matrices generated for each species indicate a generally high classification accuracy, with only a few misclassifications. The classification report confirms strong precision and recall for both presence and absence classes, with overall model accuracy exceeding 90% for all three species.

Feature Importance: To better understand the ecological drivers of invasion, we analyzed feature importance using the trained Random Forest models (Figure 6):

- ***Buddleja davidii***: The most important variable is mean temperature, followed by mean elevation at 5m and 250m scales. This suggests its preference for specific temperature ranges and elevation gradients.
- ***Prunus laurocerasus***: Total building area and mean temperature are the dominant features, highlighting its strong association with urban areas and particular climatic conditions.
- ***Robinia pseudoacacia***: Mean temperature, mean elevation (5m and 250m scales), and proximity to highways emerge as key features, reflecting its adaptability to temperature gradients and human infrastructure.

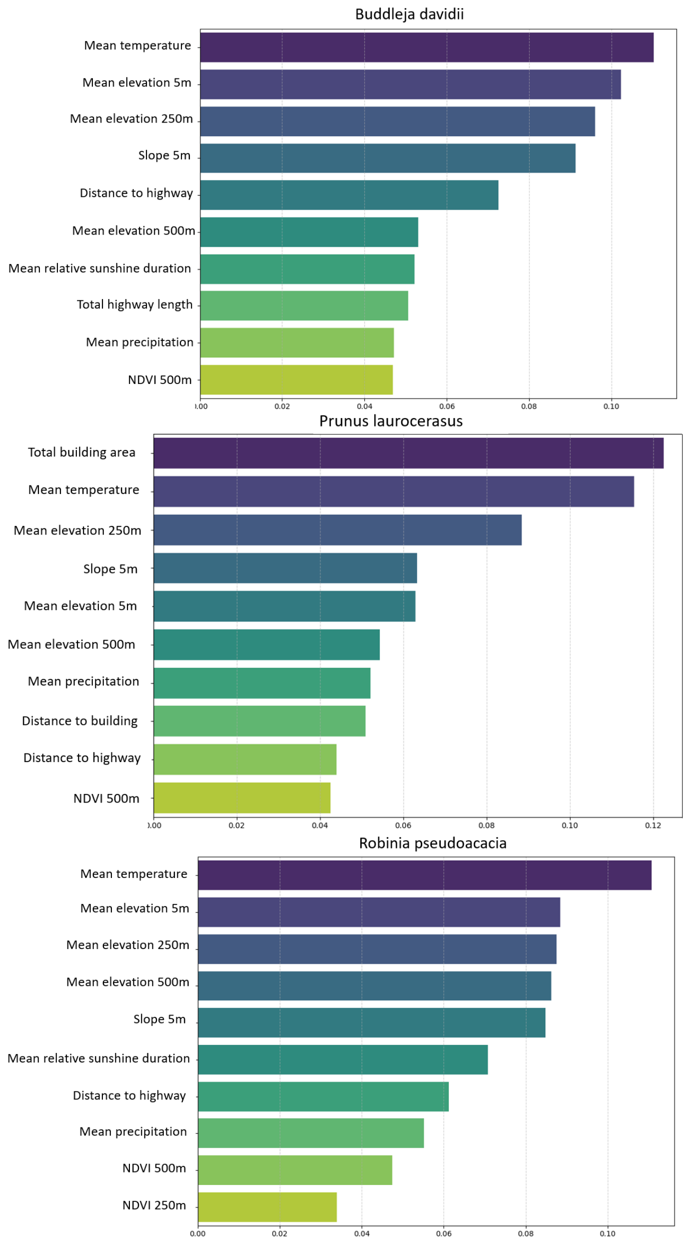

Figure 6. Top feature importances for predicting the presence of Buddleja davidii, Prunus laurocerasus, and Robinia pseudoacacia

Finally, the trained models were applied across the entire study area, producing maps that illustrate the probability of distribution for all three invasive plant species. The maps show that these species are predominantly observed in urbanized areas or near human infrastructure, such as roads. This highlights the importance of incorporating features that capture information related to human activity patterns, rather than relying solely on environmental and climatic variables.

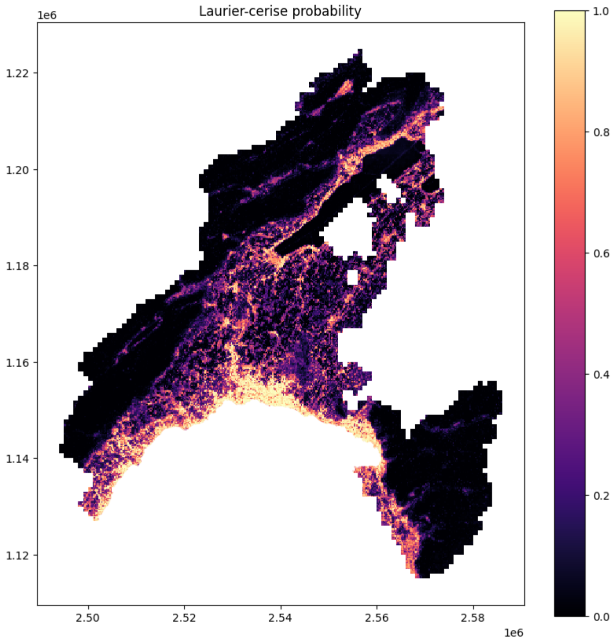

Figure 7. Example of the spatial probability map for Prunus laurocerasus

5 Discussion and conclusion

The spread of invasive alien plant species poses a significant threat to biodiversity, ecosystem stability, and human infrastructure (Kumar Rai and Singh, 2020). As these species establish and expand, they can outcompete native flora, alter habitat conditions, and impose considerable economic costs (Linders et al., 2019). Predicting their distribution patterns is a key step toward more effective management and mitigation efforts (Jarnevich et al., 2023).

In this study, we used species distribution modeling to analyze the environmental and human-related factors influencing the spread of three of the high-priority invasive species in Switzerland: Prunus laurocerasus, Buddleja davidii, and Robinia pseudoacacia. Using a dataset combining citizen science observations, land-use data, meteorological variables, and topographical features, we trained Random Forest models to predict their potential distribution. Our results highlight the dominant role of mean temperature, elevation, and proximity to human infrastructure in shaping the presence of these species.

The spatial probability maps generated in this study provide valuable insights for conservation practitioners. These maps can guide targeted intervention efforts by prioritizing high-risk areas for eradication, optimizing resource allocation, and ultimately preventing the further spread of invasive plants.

However, it is important to acknowledge that the observational data used in this study, particularly those derived from citizen science contributions, are subject to spatial sampling bias. Many species records are clustered along roads or in accessible areas, potentially influencing model outputs—especially in relation to distance-to-road variables. Future work should explore strategies to correct for this bias, such as effort-aware sampling designs, bias-corrected background selection, or integrating sampling effort covariates into the modeling process.

In addition, future research should explore dynamic modeling approaches to assess both historical trends and future expansion under different climate change and land use scenarios. Integrating additional socio-ecological data could enhance predictive accuracy and better capture the influence of human activities and landscape changes on species spread. Another crucial step is to raise public awareness and encourage citizen scientists to contribute data, particularly in underrepresented areas, to minimize observational biases and improve dataset reliability.

Acknowledgements

This study has been partially supported by the Swiss National Science Foundation (SNSF) Agora Grant No. 225132.

Declaration of Generative AI in writing

The authors utilized AI tools, such as ChatGPT and Quillbot, exclusively to enhance the language and readability of the article. They carefully reviewed the text and take full responsibility for its content.

References

Andersen, M. C., Adams, H., Hope, B., and Powell, M.: Risk Assessment for Invasive Species, Risk Analysis, 24, 787–793, https://doi.org/10.1111/j.0272-4332.2004.00478.x, 2004.

Callaghan, C. T., Winnebald, C., Smith, B., Mason, B. M., and López-Hoffman, L.: Citizen science as a valuable tool for environmental review, Frontiers in Ecology and the Environment, 23, https://doi.org/10.1002/fee.2808, 2024.

Cho, K. H., Park, J.-S., Kim, J. H., Kwon, Y. S., and Lee, D.-H.: Modeling the distribution of invasive species (Ambrosia spp.) using regression kriging and Maxent, Frontiers in Ecology and Evolution, 10, 1036 816, https://doi.org/10.3389/fevo.2022.1036816, 2022.

Elith, J. and Leathwick, J. R.: Species Distribution Models: Ecological Explanation and Prediction Across Space and Time, Annual Review of Ecology, Evolution, and Systematics, 40, 677–697, https://doi.org/10.1146/annurev.ecolsys.110308.120159, 2009.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J.: A statistical explanation of MaxEnt for ecologists: Statistical explanation of MaxEnt, Diversity and Distributions, 17, 43–57, https://doi.org/10.1111/j.1472-4642.2010.00725.x, 2010.

Elith, J., Graham, C., Valavi, R., Abegg, M., Bruce, C., Ford, A., Guisan, A., Hijmans, R. J., Huettmann, F., Lohmann, L., Loiselle, B., Moritz, C., Overton, J., Peterson, A. T., Phillips, S., Richardson, K., Williams, S., Wiser, S. K., Wohlgemuth, T., and Zimmermann, N. E.: Presence-only and Presence-absence Data for Comparing Species Distribution Modeling Methods, Biodiversity Informatics, 15, 69–80, https://doi.org/10.17161/bi.v15i2.13384, 2020.

European Commission DG Environment: Invasive plants: new study indicates how to prioritise species for management, https://environment.ec.europa.eu/news_en, [Accessed 03-02-2025], 2024.

Federal Office for the Environment (FOEN), ed.: Alien Species in Switzerland. An inventory of alien species and their impact, no. 2220 in Environmental studies, Federal Office for the Environment, Bern, 1st updated edition 2022, 1st edition 2006 edn., 2022.

Fox, E. W., Hill, R. A., Leibowitz, S. G., Olsen, A. R., Thornbrugh, D. J., and Weber, M. H.: Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology, Environmental Monitoring and Assessment, 189, https://doi.org/10.1007/s10661-017-6025-0, 2017.

Grinsztajn, L., Oyallon, E., and Varoquaux, G.: Why do tree-based models still outperform deep learning on tabular data?, https://doi.org/10.48550/ARXIV.2207.08815, 2022.

Info Flora: Info Flora - The National Data and Information Center for Swiss Flora, https://www.infoflora.ch/en/, accessed: 2025-02-04, 2025.

Jarnevich, C., Engelstad, P., LaRoe, J., Hays, B., Hogan, T., Jirak, J., Pearse, I., Prevéy, J., Sieracki, J., Simpson, A., Wenick, J., Young, N., and Sofaer, H. R.: Invaders at the doorstep: Using species distribution modeling to enhance invasive plant watch lists, Ecological Informatics, 75, 101 997, https://doi.org/10.1016/j.ecoinf.2023.101997, 2023.

Keller, R. P., Geist, J., Jeschke, J. M., and Kühn, I.: Invasive species in Europe: ecology, status, and policy, Environmental Sciences Europe, 23, https://doi.org/10.1186/2190-4715-23-23, 2011.

Kumar Rai, P. and Singh, J.: Invasive alien plant species: Their impact on environment, ecosystem services and human health, Ecological Indicators, 111, 106 020, https://doi.org/10.1016/j.ecolind.2019.106020, 2020.

Linders, T. E. W., Schaffner, U., Eschen, R., Abebe, A., Choge, S. K., Nigatu, L., Mbaabu, P. R., Shiferaw, H., and Allan, E.: Direct and indirect effects of invasive species: Biodiversity loss is a major mechanism by which an invasive tree affects ecosystem functioning, Journal of Ecology, 107, 2660–2672, https://doi.org/10.1111/1365-2745.13268, 2019.

Lotfian, M., Ingensand, J., and Brovelli, M. A.: AN APPROACH FOR REAL-TIME VALIDATION OF THE LOCATION OF BIODIVERSITY OBSERVATIONS CONTRIBUTED IN A CITIZEN SCIENCE PROJECT, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W1-2022, 271–278, https://doi.org/10.5194/isprs-archives-xlviii-4-w1-2022-271-2022, 2022.

Lotfian, M., Ingensand, J., and Claramunt, C.: Towards a Multidimensional Interaction Framework for Promoting Public Engagement in Citizen Science Projects, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, https://doi.org/10.4230/LIPICS.GISCIENCE.2023.8, 2023.

McSHEA, W. J.: What are the roles of species distribution models in conservation planning?, Environmental Conservation, 41, 93–96, https://doi.org/10.1017/s0376892913000581, 2014.

Miller, J.: Species Distribution Modeling, Geography Compass, 4, 490–509, https://doi.org/10.1111/j.1749-8198.2010.00351.x, 2010.

Moradi, M., Roche, S., and Mostafavi, M. A.: Exploring five indicators for the quality of OpenStreetMap road networks: a case study of Québec, Canada, Geomatica, 75, 1–31, https://doi.org/10.1139/geomat-2021-0012, 2021.

Roy, H. E., Pauchard, A., Stoett, P., Renard Truong, T., Bacher, S., Galil, B. S., Hulme, P. E., Ikeda, T., Sankaran, K., McGeoch, M. A., Meyerson, L. A., Nuñez, M. A., Ordonez, A., Rahlao, S. J., Schwindt, E., Seebens, H., Sheppard, A. W., and Vandvik, V.: IPBES Invasive Alien Species Assessment: Summary for Policymakers, https://doi.org/10.5281/ZENODO.7430692, 2024.

Simon, S. M., Glaum, P., and Valdovinos, F. S.: Interpreting random forest analysis of ecological models to move from prediction to explanation, Scientific Reports, 13, https://doi.org/10.1038/s41598-023-30313-8, 2023.

Stohlgren, T. J. and Schnase, J. L.: Risk Analysis for Biological Hazards: What We Need to Know about Invasive Species, Risk Analysis, 26, 163–173, https://doi.org/10.1111/j.1539-6924.2006.00707.x, 2005.

swisstopo: SwissALTI3D - High Precision Elevation Model of Switzerland, https://www.swisstopo.admin.ch/en/height-model-swissalti3d, accessed: 2025-02-11, 2025.