GeoTraining

Group members: Amira Ali, Cindy Madrid, Imas Masithoh Devangsari

COURSE: SPECIES DISTRIBUTION PREDICTION

OER Planning

Section 1: Introduction to the course

Title	Species Distribution Prediction
Duration	3 days. Eight hours per day.
Introduction	Welcome to our course. Along with contemporary developments and advances in technology in the fields of research and education, teachers are required to be more creative in following existing developments. We created this course to introduce a learning machine that will help teachers in the learning process. This course is targeted at professors teaching higher education with an intermediate knowledge of remote sensing data analysis. The Species Distribution Prediction is running as 3 days (8 hours per day) of training in the field of geospatial technologies. Its participants will be qualified to work with geospatial technologies and their application for environmental and remote sensing data analysis. The content and learning outcomes of each module will enable students to achieve the expected learning outcome. This course will discuss distribution prediction species. There are 3 modules in this course which contain code, data, and steps to prepare land cover classification and assessment: ● Module 1 (8h) includes an introduction to environmental data processing and remote sensing through geospatial technology. ● Module 2 (8h) contains how to make a land cover classification. ● Module 3 (8h) contains how to make a species distribution prediction. Apart from that, we also provide various literature to help understand R Studio and QGIS. Below in Table 1, you may find the content for each module, along with the learning outcomes, duration, and materials needed.

Table 1. Content of The Course

Module	Topic	Learning Outcomes	Duration	Materials
Module 1	Introduce to Remote Sensing	Understand how to integrate GIS and R studio to develop a Land Cover Classification	8 hours	Open Resource (YouTube)
Module 2	Land Cover Classification	Apply machine learning to predict the allocation of species (Natural vs Introduced) in the Fogo Island.	8 hours	R studio, QGIS
Module 3	Species Distribution Prediction	Evaluate the prediction model and establish the vicinity of species around the settlements.	8 hours	R studio, QGIS, RF model

Here, we introduce learning machines and tools such as R studio and QGIS to predict species distribution. The course will use constructive learning theory using study case problems. Contextualized knowledge is applicable knowledge for specific situations, such as Fogo Island.

Learning outcomes

After completing this course, the learners will be able to:

1. Describe machine learning such as QGIS and R studio
2. Define polygons using QGIS to develop a Land Cover Classification
3. Practice machine learning in order to predict the allocation of species (Natural vs. Introduced) in Fogo Island.
4. Measure the accuracy of the prediction model and establish the vicinity of species around the settlements.
5. Interpret data using machine learning
6. Categorize the types of species and explain the results based on the distance from settlements

Material	Materials the learner needs for working: ● QGIS-Software ● R Studio Software ● Remote Sensing Data (provided in this course) ● Open Source (YouTube)
Literature	Basic literature: ● Take a look at this book where are the first steps for those beginners to R and RStudio. Author: Christine Monnier https://oercommons.org/courses/r-and-r-studio-for-absolute-beginners-simple-book-publishing ● With this resource, you may be able to understand georeference maps. Here, you will learn how to add raster layers to GIS software: https://oercommons.org/courses/the-programming-historian-2-georeferencing-in-qgis-2-0 ● Below are links to YouTube videos for an introduction, setup, and scripts on how to use R in QGIS: https://www.youtube.com/watch?v=GyHgjcAYVW4
Metadata	Name: Course Species Distribution Prediction Learning ResourceType: Open Resource Description: Course composed of 3 modules targeted to intermediate level, especially professors to assess remote sensing data analysis via QGIS and R for prediction of species. Keyword(s): Remote Sensing Data, Land Cover Classification, Prediction Model Author(s): Amira Ali, Cindy Madrid, Imas Devangsari Language: English License: This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  Creation Date: September 08th, 2023 Link Github: https://github.com/Amira-AlSamawi/LearningUnit-speciesDistributionPrediction/blob/main/Module3_code.r

Section 2: Input:

Module 1

Welcome to our first Module: Introduction to Remote Sensing.
This module consists of two major parts:

1. Introduction to Remote Sensing
2. Concepts of Remote Sensing

Let's Start

Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance (typically from satellite or aircraft)**.** Below is a short presentation on Introduction to Remote Sensing provided by the University of Marburg and Goethe University.

To start the module you can download the slide below.
https://docs.google.com/presentation/d/1GqXQiiA7zaTjLunB7IsAVpi0XKIUPPex6Lf_Uc8QhF4/embed?start=false&loop=false&delayms=3000&slide=id.p4

After knowing about remote sensing, then everything will be about remote sensing and earth observation. As broad as this field is nowadays, we want to give you an overview of important concepts of remote sensing along with some best practices to quickly get you to work with satellite imagery. You can follow along the videos in this YouTube Playlist.

https://www.youtube.com/watch?v=Zo31paYaGrM&list=PL1MbwuMcC4yWw0UhPyJE-oJCfXLIYrBYM

Module 2: Land Cover Classification

After a full day of Remote Sensing, today everything will be about application Remote Sensing. Among the applications of remote sensing analysis, in this module, you will learn how to elaborate land cover classifications. They are useful tools for the exploration, monitoring, and follow-up of environmental effects on landscapes and the impacts of anthropogenic influences.

Also with this module, you will learn to use machine learning. Machine learning is the development of computer systems that with the help of algorithms and statistics elaborate conclusions based on patterns. The advantage of using machine learning is they adapt and are able to learn through instructions that coders develop.

Below you may find a literature review of machine learning algorithms:
https://www.researchgate.net/publication/344717762_Machine_Learning_Algorithms_-A_Review

After defining the machine learning algorithm, you will be able to predict the species around the settlement with the remote sensing data provided.

Now you are able to start working with the Land Cover Classification and drawing the polygons. Please find below the training provided to work with QGIS and use the information presented on the link of Concepts of Brief Remote Sensing Data. If you need any help with the code in R, you may find it here.
https://www.youtube.com/watch?v=C6B_kUHCmWs

Data needed to do the Land Cover Classification. Upload the TIF into the QGIS:
https://drive.google.com/drive/folders/1f9pEdYGeqvokdRINNlAZyF-eu3ixEY3I

Module 3

As the Land Cover Classification is done. We want to encourage you to work with the data below. You may find the Excel data in order to correlate the species with the settlements. You may find help with the code here

Data:
https://drive.google.com/drive/folders/1f9pEdYGeqvokdRINNlAZyF-eu3ixEY3I

So far, we have the Land Cover Classification imported into R and the correlation between the species and the settlements. Now is the time to work with species distribution modeling. Below is a GitHub with information about species distribution modeling using R:

https://damariszurell.github.io/SDM-Intro/

We hope with this GitHub you may be able to establish a model. If further help is needed with the coding please follow this link.

Now we encourage you to determine a way to predict whether natural or introduced species are near or far from settlements. Hint: You can use the link below with the coding.

We wish you all the best and hope you can enjoy and succeed in every module of this course!

Section 3: Application

After an introduction to R Studio, we move on to remote sensing and earth observation. Below are the links to an overview of applications on remote sensing with some practices including imagery satellite.

Application of Remote sensing data includes lectures and R practice:
https://geotraining.geomedienlabor.de/doku.php?id=en:courses:training:element-01:worksheets:lc-ws-03-1

For further understanding. Please follow the link to know the Current applications on remote sensing:
https://www.youtube.com/watch?v=lwdcda8lbbg&list=PL1MbwuMcC4yWw0UhPyJE-oJCfXLIYrBYM&index=3

Section 4: Assessment and Wrap up

The following assessment criteria are adopted to evaluate the fulfillment of the learning outcomes:

Assessment Aspect	Percentage (100%)
Accuracy	20%
Quality of the polygons	30%
Understanding	30%
Practical skills	20%

Accuracy

The accuracy can be measured from the confusion matrix using the following equation:

ACC =number of samples in the diagonal / number of all samples.

The accuracy enhances as it approaches the value "1". Refer to the code of module 3 for the implementation using R

Quality of the polygons

As the accuracy highly depends on the polygons, and high accuracy might be misleading if there were only a few polygons or they were chosen from only a few classes, the quality of the polygons must be considered. For this, the following questions must be answered to assess the quality of the polygons:

1- How many polygons are constructed? A reasonable number would be between 20 and 80

2- Are the polygons equally distributed between different land cover classes(number and size of polygons for each class)? There must be sufficient polygons from each class.

3- What is the range of the size of the polygons? Too small polygons will not contain sufficient sample points, and too big polygons might be noisy. The polygons on the ocean might be big as no other classes are expected in that region.

Uderstanding

A discussion with students might be conducted to evaluate their understanding. The following questions might be asked during the discussion:

1- What is the effect of the radius size around the settlement?

2- If you have a low classification accuracy, what are some possible ways to enhance it?

Practical skills

Students gained skills can be evaluated in terms of their ability to use platforms, train models, and write clean code.

Dimension of performance	Not acceptable	Pass	Satisfactory	Highly Satisfactory	Best expectable outcome
R studio					
QGIS					
RF model					

Section 5: Appendix

Scientific reason:

- Monnier, Christine. R and R Studio for absolute beginners. College of DuPage Press, 2022.
- Mahesh, Batta. (2019). Machine Learning Algorithms -A Review. 10.21275/ART20203995.
- Barbet-Massin M, Rome Q, Villemant C, Courchamp F. Can species distribution models really predict the expansion of invasive species? PLoS One. 2018 Mar 6;13(3):e0193085. doi: 10.1371/journal.pone.0193085. PMID: 29509789; PMCID: PMC5839551.
- Valavi, Roozbeh., Guillera-Arroita, Gurutzeta., Lahoz-Monfort, Jose J., et al. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs, vol. 92, no. 1. February 2022, doi: 10.1002/ecm.1486.
- Zhang, C., Chen, Y., Xu, B. et al. Improving prediction of rare species' distribution from community data. Sci Rep 10, 12230 (2020). doi: 10.1038/s41598-020-69157-x.
- Sara Beery, Elijah Cole, Joseph Parker, Pietro Perona, and Kevin Winner. 2021. Species Distribution Modeling for Machine Learning Practitioners: A Review. In ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '21), June 28-July 2, 2021, Virtual Event, Australia. ACM, New York, NY, USA 20 Pages. doi: 10.1145/3460112.3471966.